

Query Match	7.28;	Score 72.5;	DB 4;	Length 284;
Best Local Similarity	23.88;	Pred. No. 1.3;		

Storage  
88-11-27-05  
Filed



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 13:54:32 ; Search time 88 Seconds  
(without alignments)  
379.409 Million cell updates/sec

Title: US-09-881-636-2  
Perfect score: 1002  
Sequence: 1 MVAATGSLSKNKPASISELD.....FRLVKKLYSLIGTVIEGS 193

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1002	100.0	193	4	096D03 homo sapien
2	932	93.0	193	11	09CQV2 09CQV2 mus musculu
3	271	27.0	232	4	09H0S3 09H0S3 homo sapien
4	271	27.0	232	4	09NX09 09NX09 homo sapien
5	262	26.1	229	11	09D3F7 09D3F7 mus musculu
6	211.5	21.1	299	5	09NHNS 09NHNS drosophila
7	211.5	21.1	299	5	09VT18 09VT18 drosophila
8	193	19.3	280	5	09VTH4 09VTH4 drosophila
9	188	18.8	280	5	09NH44 09NH44 drosophila
10	91	9.1	902	12	09OB81 09OB81 yaba monkey
11	86.5	8.6	1089	10	09FN97 09FN97 arabidopsis
12	85.5	8.5	1414	5	076411 076411 caenorhabdi
13	84	8.4	902	12	09DHL2 09DHL2 yaba-like d
14	82.5	8.2	2081	10	09LH98 09LH98 arabidopsis
15	82	8.2	673	5	076649 076649 caenorhabdi
16	82	8.2	1002	16	097F23 097F23 clostridium

17	81.5	8.1	228	17	097YG5 097YG5 sulfolobus
18	80.5	8.0	498	4	09NPE8 09NPE8 homo sapien
19	80.5	8.0	582	4	09NSP8 09NSP8 homo sapien
20	80	8.0	443	5	09U0X3 09U0X3 letismania
21	80	8.0	927	4	075170 075170 homo sapien
22	79.5	7.9	606	5	044940 044940 drosophila
23	79.5	7.9	1704	5	095206 095206 trypanosoma
24	79	7.9	1787	10	09C507 09C507 arabidopsis
25	78.5	7.8	680	5	015740 015740 dictyosteli
26	78	7.8	249	2	093UD4 093UD4 candidatus
27	78	7.8	315	11	093UV5 093UV5 mus musculu
28	78	7.8	1134	12	099D22 099D22 bovine herp
29	77.5	7.7	817	10	092P69 092P69 arabidopsis
30	77.5	7.7	1153	10	09SL22 09SL22 arabidopsis
31	77.5	7.7	1584	10	09MAG6 09MAG6 arabidopsis
32	77	7.7	600	4	096MN7 096MN7 homo sapien
33	77	7.7	674	11	09DCX9 09DCX9 mus musculu
34	77	7.7	1095	10	09LU25 09LU25 arabidopsis
35	77	7.7	1837	10	09LU53 09LU53 arabidopsis
36	76.5	7.6	574	5	0960X1 0960X1 drosophila
37	76.5	7.6	840	5	09N3Q5 09N3Q5 caenorhabdi
38	76.5	7.6	1262	5	09W110 09W110 drosophila
39	76.5	7.6	1309	10	09SXF0 09SXF0 arabidopsis
40	76.5	7.6	1466	10	094I09 094I09 arabidopsis
41	76	7.6	332	10	094LH5 094LH5 oryza sativ
42	76	7.6	1083	3	096I08 096I08 saccharomyc
43	76	7.6	1871	10	09CAG6 09CAG6 arabidopsis
44	76	7.6	1894	10	064795 064795 arabidopsis
45	75.5	7.5	423	16	09CNX7 09CNX7 pasteurella

ALIGNMENTS

RESULT 1

ID 096D03 PRELIMINARY: PRT: 193 AA.  
AC 096D03:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SIMILAR TO RIKEN CDNA 1700037B15 GENE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG CARCINOMA;  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC013592; AAH13592.1; -  
SQ SEQUENCE 193 AA; 21740 MW; 3BEB3D0530C5BFF CRC64;

Query Match	Score	100.0%	Score	1002	DB 4	Length	193		
Best Local Similarity	100.0%	Pred. No.	3.8e-96						
Matches	193	Conservative	0	Mismatches	0	Indels	0	Gaps	0
OY	1	MVAATGSLSKKNPASISELDDCGHPESLSDDPDYMDYVPEPNLNEVIFEESTCONLYKM	60						
Db	1	MVAATGSLSKKNPASISELDDCGHPESLSDDPDYMDYVPEPNLNEVIFEESTCONLYKM	60						
OY	61	LENCLSKSKQTKLGCSSKVLVPEKLTQRIADVLRLSTFPCGLRCQVMHVNLEINVCCK	120						
Db	61	LENCLSKSKQTKLGCSSKVLVPEKLTQRIADVLRLSTFPCGLRCQVMHVNLEINVCCK	120						
OY	121	IDRIYCDSSVPTFELTLVFKQENCSWTSFRPFSSRGFSGFRRTJLSSGFLVKKK	180						
Db	121	IDRIYCDSSVPTFELTLVFKQENCSWTSFRPFSSRGFSGFRRTJLSSGFLVKKK	180						
OY	181	LYSLIGTVIEGS 193							
Db	181	LYSLIGTVIEGS 193							

## RESULT 2

09COV2 ID 09COV2 PRELIMINARY: PRT: 193 AA.

AC 09COV2: 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DE 1700108M02RIK PROTEIN (1700037B15RIK PROTEIN).

GN 1700108M02RIK OR 1700037B15RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCB1\_TaxID=10090;

RA NCB1\_TaxID=10090;

RC STRAIN=C57BL/6J; TISSUE=TESTIS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Flieschmann W., Gaasterland T., Gissi C., King B., Kochiya H., Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J., Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons F., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.,

RA "Functional annotation of a full-length mouse cDNA collection."

RT Nature 409:685-690(2001).

RL EMBL; AK018944; BAB24490.1; -

DR EMBL; AK006617; BAB24676.1; -

DR MGD; MGI:1925834; 1700037B15RIK.

DR MGD; MGI:1925887; 1700108M02RIK.

SQ SEQUENCE 193 AA: 21597 MW: 536480776C1E5B0D CRC64;

Query Match 93.0%; Score 932; DB 11; Length 193;

Best Local Similarity 93.2%; Pred. No. 7e-89;

Matches 178; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 MWATGSLSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKK 60

DB 1 MWATGSLSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKK 60

QY 61 LENCLSKSKQTKGSKVLPKLTQRIADYRLSTFPCGRCGVMHVNLEIENVCCK 120

DB 61 LENCLSKSKQTKGSKVLPKLTQRIADYRLSTFPCGRCGVMHVNLEIENVCCK 120

QY 121 LDRIVCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKK 180

DB 121 LDRIVCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKK 180

QY 121 LDRIVCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKK 180

DB 121 LDRIVCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKK 180

QY 181 LKSLGTIVIE 191

DB 181 LKSLGTIVIE 191

RESULT 3

Q9H0S3 ID Q9H0S3 PRELIMINARY: PRT: 232 AA.

AC Q9H0S3: 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE HYPOTHETICAL 25.4 KDA PROTEIN.

GN DKFZP564O2071.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCB1\_TaxID=9606;

RA NCB1\_TaxID=9606;

RC TISSUE=TESTIS;

RX MEDLINE=21154917; PubMed=11230166;

RA Wiemann S., Wellenreuther R., Gassenhuber J., Glassl S., Ansoorge W., Boecker M., Bloeker H., Bauersachs S., Blum H., Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N., Mewes H.W., Korn B., Klein M., Poustka A., Wambutt R., Korn B., Klein M., Poustka A., "Towards a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs."

RT Genome Res. 11:422-435(2001).

RL EMBL; AL136668; CAB66603.1; -

DR EMBL; AL136668; CAB66603.1; -

SQ SEQUENCE 232 AA: 25354 MW: 774E941EAD908198 CRC64;

Query Match 27.0%; Score 271; DB 4; Length 232;

Best Local Similarity 35.0%; Pred. No. 3.9e-20;

Matches 63; Conservative 45; Mismatches 66; Indels 6; Gaps 4;

QY 6 SLSSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 64

DB 48 SLSSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 64

QY 65 LSKSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 124

DB 104 LSKSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 124

QY 125 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

DB 164 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

QY 164 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

DB 164 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

RESULT 4

Q9H0S3 ID Q9H0S3 PRELIMINARY: PRT: 232 AA.

AC Q9H0S3: 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE CDNA FLJ20500 FIS; CLONE KAT09159 (HYPOTHETICAL 25.4 KDA PROTEIN).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCB1\_TaxID=9606;

RA NCB1\_TaxID=9606;

RC TISSUE=TESTIS;

RX MEDLINE=21154917; PubMed=11230166;

RA Wiemann S., Wellenreuther R., Gassenhuber J., Glassl S., Ansoorge W., Boecker M., Bloeker H., Bauersachs S., Blum H., Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N., Mewes H.W., Korn B., Klein M., Poustka A., Wambutt R., Korn B., Klein M., Poustka A., "Towards a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs."

RT Genome Res. 11:422-435(2001).

RL EMBL; AL136668; CAB66603.1; -

DR EMBL; AL136668; CAB66603.1; -

SQ SEQUENCE 232 AA: 25354 MW: 774E941EAD908198 CRC64;

Query Match 27.0%; Score 271; DB 4; Length 232;

Best Local Similarity 35.0%; Pred. No. 3.9e-20;

Matches 63; Conservative 45; Mismatches 66; Indels 6; Gaps 4;

QY 6 SLSSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 64

DB 48 SLSSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 64

QY 65 LSKSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 124

DB 104 LSKSKNPASISSELLDCGYHPESLLSDPDYNDYVPEPRLNVIYEESTCQNLVKKLENC 124

QY 125 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

DB 164 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

QY 164 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

DB 164 VCDSSVVPFELTLVFKQENCSWTSFRDFEFSRGFRSSGFRRLILSSGFRVLVKKL 183

SO SEQUENCE 232 AA; 25370 MW; 774E941EBDD08198 CRC64;

Query Match 27.0%; Score 271; DB 4; Length 232;  
Best Local Similarity 35.0%; Pred. No. 3.9e-20;  
Matches 63; Conservative 45; Mismatches 66; Indels 6; Gaps 4;

QY 6 SLSSKNPASIELDCCGYHPESILSDPDYMDYV-VPEPNLMEVIFEESTCONLVKMLENC 64  
DB 48 SLESSDCSL-DSSNSGCGPE---EDTAYLDGVSLLPDELLSDPDEHLCAIMQLQES 103  
QY 65 LSKSKOTKLGCSKVLPEKLTQRIADYLRSTPECGLRGCMHVNIEIENVCCKLRI 124  
DB 104 LAQARLGSRPARRLMPSQVSGKELRLAYSPGCLRGALLDVCEGKSCHSVQL 163  
QY 125 VCDSSVVFELTLVFKOENCSWTSFRDFEFS-RCGESSGFRRLYIISGFLVKKLYS 183  
DB 164 ALDPSLVPTFQTLTLRLDRSLRMPKIQGLFSSANSPLFGFSOSLTLTGFRVYKKLYS 223

RESULT 5  
Q9DJF7 PRELIMINARY; PRT; 229 AA.

AC Q9DJF7;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 5830413E08RIK PROTEIN.  
GN 5830413E08RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=THYMUS;  
RA MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aizawa K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochava H.,  
RA Knehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,  
RA Blake J., Boffelli D., Bojunga N., Carmucci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilmink L.,  
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL; AK017926; BAB31006.1;  
DR MGI; MGI:1921997; 5830413E08RIK.  
SQ SEQUENCE 229 AA; 24871 MW; FE7EA24D3FE63949 CRC64;

Query Match 26.1%; Score 262; DB 11; Length 229;  
Best Local Similarity 30.7%; Pred. No. 3.3e-19;  
Matches 65; Conservative 45; Mismatches 68; Indels 34; Gaps 5;

QY 3 ATGSLSSKNPA-----SISELLDC-----GYPEISLSPD 33  
DB 12 SSSSSSTTPADRPSPSAMGSAAREGLDRCALESSDCSLSSNSGCGPE---EDSS 68  
QY 34 YMDYV-VPEPNLMEVIFEESTCONLVKMLENCSKOTKLGCSKVLPEKLTQRIADY 92  
DB 69 YLDGVSLLPDELLSDPDEHLCAIMQLQESLQARLGSRPARRLMPSQVSGKEL 128  
QY 93 LRLSTPECGLRGCMHVNIEIENVCCKLRIVCDSSVVFTELTLVFKOENCSWTSFRD 152

DB 129 LRLAYSEPCGLRGALDLVCEGKSCHSVAQALDPSLVPTFQTLTLRLDRSLMPTIOG 188  
QY 153 FEFS-RCGESSGFRRLYIISGFLVKKLYS 183  
DB 189 LLSANSSILVPCYSOSLTLTGFRVYKKLYS 220

RESULT 6  
Q9NHN5 PRELIMINARY; PRT; 299 AA.

AC Q9NHN5;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CHARYBDE.  
GN CHARYBDE OR CG7533.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chauvet S., Maurel-Zaffran C., Miassod R., Julien N., Pradel J.,  
RA Aragnol D.;  
RT "Characterization of charybde and scylla, two paralogous target genes  
of hox and cofactors proteins in Drosophila.";  
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF221109; AAF59840.1;  
DR FlyBase; FBgn0036165; charybde.  
SQ SEQUENCE 299 AA; 32133 MW; 9D6560B8963315E0 CRC64;

Query Match 21.1%; Score 211.5; DB 5; Length 299;  
Best Local Similarity 35.9%; Pred. No. 8e-14;  
Matches 51; Conservative 24; Mismatches 60; Indels 7; Gaps 3;

QY 43 NLNVEIFEESTCONLVKMLENCSKOTKLGCSKVLPEKLTQRIADYLRSTPECG 102  
DB 162 NLDDV-SASAVRELSQLQALRDARHRLACTEVLTPNDLORIAEIIIRMSREPCG 219  
QY 103 LRGCMVHNLETE-NVCKRLRIYCDSSVVFTELTLVFKOENCSWTSFRDFEFSRGFS 161  
DB 220 ERACLTLEFESEPPKVKRIAFKDPDTVSIFELYLRLRODKSSSLVPOFIKNTLS 279  
QY 162 SGFRRLYIISGFLVKKLYS 183  
DB 280 N---TINISPDFTFKKKLYS 297

RESULT 7  
Q9VTI8 PRELIMINARY; PRT; 299 AA.

AC Q9VTI8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CHARYBDE PROTEIN (LD22381P).  
GN CHARYBDE OR CG7533.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Morten J.R., Yandell M.D., Zhang O., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abill J.F., Abgayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
RA Bailett R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman J.D., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster."  
RT Science 287:2185-2195(2000).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN=, CN BW SP:  
RA Stapleton M., Brokstein P., Hong L., Abgayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,  
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE005544; AAF50060.2; -  
DR FlyBase: FBgn0036165; charybde.  
SQ FlyBase: FBgn0036165; charybde.  
SQ SEQUENCE 299 AA; 32179 MW; CD224D3BFBF315E7 CRC64;  
  
Query Match 21.1%; Score 211.5; DB 5; Length 299;  
Best local Similarity 35.9%; Pred. No. 8e-14;  
Matches 51; Conservative 24; Mismatches 60; Indels 7; Gaps 3;  
  
QY 43 NINEVFEESTCONLVKMLENCLSKSQTKLGGSKVLVPEKLTQRIADVLRISTEPCG 102  
DB 162 NIDDDV--SASAVRELSQLQAOQLRDAKRRLACTEVLTPNDLQRIAAELITRSEPCG 219  
QY 103 LRGCVAHVNLIEF--NCKKIDRLIVCDSSVVPFELTVKQENCSWTFSPDFFSKGRS 161  
DB 220 ERACLTFFIEESEPKNKRAVFKVPDPVTSIFELYLTLLRQDKSGWSLVPQIKNLTRS 279  
QY 162 SGRFRTLILSSGFRVKKLYS 183  
DB 280 N-----TINISPDFTLTKKLYS 297  
  
RESULT 8  
Q9VTH4 PRELIMINARY; PRT; 280 AA.  
AC Q9VTH4:  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CG7590 PROTEIN.  
GN SCYLIA OR CG7590.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abill J.F., Abgayani A., An H.-J., Andrews-Pfankuch L., Beasley E.M.,  
RA Bailett R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster."  
RT Science 287:2185-2195(2000).  
DR EMBL: AE003545; AAF50074.1; -  
DR FlyBase: FBgn0041094; scyllia.  
SQ SEQUENCE 280 AA; 30762 MW; 894841884CD28258 CRC64;  
  
Query Match 19.3%; Score 193; DB 5; Length 280;  
Best local Similarity 29.2%; Pred. No. 6.1e-12;  
Matches 49; Conservative 36; Mismatches 61; Indels 22; Gaps 5;  
  
QY 30 SDFDYMDYVPE-----PNLNEVFEESTCONLVKMLENCLSKSQTKLGGSKVL 79  
DB 99 SNBNYYYAADEEGSGADYALSNYDKAVEELS---LRILDE--LRAAKSRRLATTEVS 153  
QY 80 VPEKLTQRIADVLRISTEPCGLRGCVAHVNLIEF--NCKKIDRLIVCDSSVVPFELTL 188  
DB 154 LPCLDTPSVAREITIRVSEKPRGIRGCTIYIEFEDEKKSRRSLASTKIVSDTVSTREYVL 213  
QY 139 VFQKQENCSWTFSPDFFSKGRFRTLILSSGFRVKKLYSL 186  
DB 214 TLRQDHRGWTSLP-----QFMKSLARTITTSPEYTIIRKNKLYSDG 255  
  
RESULT 9  
Q9NHN4 PRELIMINARY; PRT; 280 AA.  
AC Q9NHN4:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)

```

DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, last annotation update)
DE SCYLLA.
GN SCYLLA OR CG7590.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Chauvet S., Maurel-Zafran C., Miasod R., Julien N., Pradel J.,
RA Aragnol D.; "Characterization of charybde and scylla, two paralogous target genes
RT of Hox and cofactors proteins in Drosophila.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF221110; AAF59841.1; -
DR FLYBASE; FBgn0041094; scylla.
SQ SEQUENCE 280 AA; 30772 MW; 8B27BC184CD14E97 CRC64;

Query Match 18.8%; Score 188; DB 5; Length 280;
Best Local Similarity 28.6%; Pred. No. 2e-11;
Matches 48; Conservative 36; Mismatches 62; Indels 22; Gaps 5;

OY 30 SDFDWDVPE-----PNIIEVFESTQNLVYKMLENLSKQTKLGGSKVL 79
DB 99 SNSNYVYAADEEGSGADVALSNYDKRAVELS---LRLDE-LRAKSHMLCTEVS 153
OY 80 VPEKLTORIADVLRLSTFPCGLRGVYHNLIEI-NVCKLRIRIVCDSSVPPFELTL 138
DB 154 LPCDLTPVAREIIRVSKFPGRCIGCTIYIEFEDEPKNSRIASIKVDPDTVSTFEVYL 213
OY 139 VFKEQNCWTSFRDFFSRGPRSSGFRRLTILSSGFRLLVKKIYSLIG 186
DB 214 TLRQDHRGWTSLP-----QFMKSLARTITISPEYITTKNLYSADG 255

RESULT 10
OY 090881 PRELIMINARY; PRT; 902 AA.
AC 090881;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, last annotation update)
DE YB-B17L.
GN YB-B17L.
OS Yaba monkey tumor virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Yatapoxvirus.
OX NCBI_TaxID=38804;
RN [1]
RP SEQUENCE FROM N.A.
RA Amano H., Morikawa S., Ueda Y., Miyamura T.;
RT "Nucleotide sequence of the central 50kbp region of Yaba virus DNA.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95248279; PubMed=7730796;
RA Amano H., Ueda Y., Miyamura T.;
RT "Identification and characterization of the thymidine kinase gene of
RT Yaba virus.";
RL J. Gen. Virol. 76:1109-1115(1995).
DR EMBL; AB015885; BAA88814.1; -
SQ SEQUENCE 902 AA; 102862 MW; ED4373030A2BCF31 CRC64;

Query Match 9.1%; Score 91; DB 12; Length 902;
Best Local Similarity 21.6%; Pred. No. 1;
Matches 50; Conservative 41; Mismatches 73; Indels 68; Gaps 11;

OY 1 MWATGSLSSKNPASISLDCGYHESLSDFDYDVVPEPNIIEVFESTC----- 54
DB 354 MIAKSKRKEKLSKLSLWD-----GMDYOET--ROKKSLDILFYNSTCYWGLY 401

```

```

OY 55 --QNLV--KMLENLSKQTKLGC-----SKVLPEKLTORI-----AQDVLRLSS 97
DB 402 NKNNTVCSMLSDIMSANDEIPLRVGVLPRVYSGKTVDPDIATLNSINSISKDDPKSS 461
OY 98 TEPCLRGCVYHNLIEIENVC--KLDRIYCDSSVPPFELTLV----- 140
DB 462 SAP-----MHIGSENNFMKFFOLLRLVMTSPERAIKELVIMFYAGIKLNDEGSPHLI 514
OY 141 KOENCWTSFRDFFSRGPRSSGFRRLTILSSG-----RLVKKIYSLI 185
DB 515 KKE---SYODFSVLFSMGFKVSIKKSIIASNNHFTIIVRPVTKOYITNML 564

RESULT 11
OY 09FN97 PRELIMINARY; PRT; 1089 AA.
AC 09FN97;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE TRANSPOSON PROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Kocani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006705; BAB09502.1; -
DR InterPro; IPR000150; Hypothet_cof.
SQ SEQUENCE 1089 AA; 126629 MW; 17664F2489EB14E CRC64;

Query Match 8.6%; Score 86.5; DB 10; Length 1089;
Best Local Similarity 19.2%; Pred. No. 3.7;
Matches 41; Conservative 33; Mismatches 83; Indels 57; Gaps 6;

OY 2 VATGSLSSKNPASISLDCG-----YHPESLSDFDYDVVPEPNIIEV 47
DB 506 ILTGELYNEVCCLPKRTVDCGCGHGRLEBGYKTHMMHKOSILMELSYMDLKRNLDMV 565
OY 48 IFEESTQNLVYKMLENLSKQTKLGGSKVLVPEKLTORIADVLRLSTFPCGLRGCV 107
DB 566 HIEKNVLDNFTLTLLNVGKTRDNIS-----RLDQENCRKD-- 604
OY 108 MHVNLIEIENVCCKLDRIYCDSSVPPFELTLVFKOENCWTSFRDFFSRGPRSS----- 162
DB 605 LHLTPE-----GKAPIKFRLKPAKEIFLRKLE-KVYKESDGYSSSLANCY 650
OY 163 ---GFRRLIILSSGFRLLVKKIKKIYSLIGTVIEGS 193
DB 651 DLPGRKLTGMKSHDCHVLMQRLPFAFALMDKS 684

RESULT 12
OY 076411 PRELIMINARY; PRT; 1414 AA.
AC 076411;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE HYPOPHETICAL 160.1 KDA PROTEIN.
GN T10H9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Greco T., Bradshaw H., O'Brien D.;
RT "The sequence of C. elegans cosmid T10H9.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF067949; AAC19236.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002172; Ldg_recept_A.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00069; pkinase; 1.
DR SMART; SM00192; LDla; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS50068; LDla_2; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Glycoprotein; Hypothetical protein; Transferase;
KW Tyrosine-protein kinase.
SQ
SEQUENCE 1414 AA; 160101 MW; 43800DFE3D48ACD4 CRC64;

Query Match
Best Local Similarity 23.6%; Score 85.5; DB 5; Length 1414;
Matches 46; Conservative 35; Mismatches 73; Indels 41; Gaps 10;

QY 16 SELLDGCGYIPESILSDP---DY---WDYVPEPNLNEYFEESTCONLVKM----- 60
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 358 SEMDCG-NINGTMCDFNGQDCNSWYQVNTVDYERLSEPTTVAPLNKLNEYVPLHFR 416
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 LENCLSKSKOTKLGCSKVLY---PEKLTQRIADVL-RLSSTFPCG-----LGCYV 108
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 417 LQSPSAKIEAMRGSGNMVLFTHKPNPLRTSALVSPELPRTNPENAYDEKSPLEFSCKL 476
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 HYNL-----ELENVCKKLDRIVCDSSVPTFELTLVFKOENCSWTSFRDFEFSRGR 159
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 477 RFLGSRYSKYWQIVISVIGKINPMESGRITIEAGYTLLPK-ENCTW---ERVFVNIPR 532
QY 160 FSSGFRRTLLISGF 174
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 533 ONAGFRIGIFVTNYF 547
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
Q9DHL2 PRELIMINARY; PRT; 902 AA.
AC Q9DHL2;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE 101L PROTEIN.
GN 101L.
OS Yaba-like disease virus (YLDV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Yatapoxvirus.
OC Yatapoxvirus.
OX NCBI_TaxID=132475;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Lee H.J.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21176366; PubMed=11277691;
RA Lee H.J., Essani K., Smith G.L.;
RT "The genome sequence of Yaba-like disease virus, a yatapoxvirus.";
RL Virology 281:170-192(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Lee H.J.;
RL Thesis (2000), Sir William Dunn School of Pathology, University of.
DR EMBL; AJ293568; CAC21339.1; -.
SQ SEQUENCE 902 AA; 103042 MW; FA106217CDDCE05 CRC64;

Query Match
Best Local Similarity 20.0%; Score 84; DB 12; Length 902;
Matches 47; Conservative 44; Mismatches 70; Indels 74; Gaps 11;

QY 1 MVAATSLSKNPASISFLDCGYHPESILSDPYWMDVPEPNLNEYFEESTC----- 54
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 354 MLAKSKSKELKSNLSLWD-----CIDQET--KQKLSLDTIFNSTCIYWGVL 401
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 55 --QNLV--KMLENCLSKSKQTKLGCSKVLYPEKLTQRIADVL-----R 94
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 402 NKNNTYCSMLSDIISFNETPLRVC---LLPRVSGKTVPDIIAETLNSINSISRKDFK 458
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 LSTBPCGLGCVMAHNLLENVCK--KLDRIYCDSSVPTFELTLV----- 140
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 459 KSAVSP-----MHIGLTENNPMKEFQQLRLVTNTPPERAIRKIVLFGALGLNDGSP 511
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 141 ---KOENCSWTSFRDFEFSRGRSSGFRRTLLISGF-----RLVKKLYSLI 185
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 512 HLIRKE--STQDPSVLLFSAMGFKVSIKSIASNNHTFIVPRVTKQIYIMWL 564
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20363099; PubMed=10907653;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety p1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002057; BAB03174.1; -.
SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFFE739 CRC64;

Query Match
Best Local Similarity 29.7%; Score 82.5; DB 10; Length 2081;
Matches 30; Conservative 13; Mismatches 29; Indels 29; Gaps 5;

QY 112 LEIENVC-----KLDRIYCDSSVPTFELTLVFKOENCSWTSFRD 152

```

```

Db      51 LKLENNCPREQLRSFFPEKLDLKLKLESSVPLIEDNEPKTFKPLKSKSEMLQTM--- 107
OY      153 FFFSGRPFSSGFRRLILSSGFRLYKK--KLXSLIGTVIE 191
      108 FMIGRGLSSSRKEM-----FEVWKSLELHAIGRVIIIE 143

```

## RESULT 15

```

O76649      PRELIMINARY;      PRT:      673 AA.
ID 076649
AC 076649:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HYPOTHETICAL 77.1 KDA PROTEIN.
GN F25E5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderiinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilson R., Bradshaw H.;
RT "The sequence of C. elegans cosmid F25E5.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF078157; AAC24078.1; -.
DR InterPro: IPR000210; TRYB_POZ.
DR InterPro: IPR001254; TRYPSIN.
DR SMART: SM00225; TRYB_1.
DR SMART: SM00020; TRYB_SPE; 1.
DR PROSITE: PS50097; TRYB; 1.
KW Hydrolase; Hypothetical protein; Serine protease.
SQ SEQUENCE 673 AA; 77122 MW; 8ED29BE157C9D8C7 CRC64;

```

## Query Match

```

      8.2%; Score 82; DB 5; Length 673;
Best Local Similarity 26.9%; Pred. No. 6;
Matches 29; Conservative 21; Mismatches 36; Indels 22; Gaps 5;

```

```

OY      2 VATGSLSKKNPASISEL-----LDCGYHPESLSDPDYWDYVPEPNLN--EVIPEE 51
      185 LARGRLNTADPYILEDITPIPFKLYMDLTYHPKQYFSAYHAKDIIQIAKRNNELEV--- 241
OY      52 STCONLV--KMLENCLSKSKQTKLGCS-----KVLVEKLTQRIAQ 90
      242 TTCQNVILEDLENKNTKNTKLAESYDLDTVRELFPNPMTCPSAR 289
Db

```

Search completed: October 11, 2002, 14:13:46  
 Job time : 92 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 11, 2002, 14:11:18 ; Search time 32 Seconds

(without alignments)  
147.317 Million cell updates/sec

Title: US-09-881-636-2

Perfect score: 1002  
Sequence: 1 MVAATGSLSSKNPASISELD.....FRLVKKKLYSLIGTTVEGS 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCrUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	8.4	286	2	US-08-576-626A-42
2	79	7.9	549	5	PCT-US93-05701-2
3	77	7.7	1417	2	US-08-559-303B-78
4	77	7.7	1417	3	US-08-781-891-78
5	77	7.7	1417	4	US-09-175-828-78
6	74	7.4	650	1	US-08-325-071-59
7	74	7.4	650	1	US-08-461-004A-59
8	72.5	7.2	284	2	US-08-767-096-3
9	72.5	7.2	284	4	US-09-480-203-3
10	72	7.2	756	4	US-09-085-199B-9
11	72	7.2	871	2	US-08-775-009-34
12	70.5	7.0	530	4	US-09-180-852-2
13	70.5	7.0	855	4	US-08-880-865A-10
14	70	7.0	650	1	US-08-325-071-56
15	70	7.0	650	4	US-08-461-004A-56
16	70	7.0	688	1	US-08-325-071-57
17	70	7.0	688	4	US-08-461-004A-57
18	70	7.0	858	3	US-08-946-026-3
19	70	7.0	1098	1	US-07-777-715-7
20	70	7.0	1098	1	US-08-170-126-2
21	70	7.0	1098	3	US-08-954-418-2
22	69.5	6.9	905	4	US-09-360-186-3
23	69	6.9	450	4	US-09-592-891A-14
24	69	6.9	549	1	US-08-325-071-61
25	69	6.9	549	1	US-08-461-004A-61
26	69	6.9	650	1	US-08-325-071-63
27	69	6.9	650	4	US-08-461-004A-63

28	69	6.9	739	2	US-08-836-943-2	Sequence 2, Appl
29	68	6.8	280	1	US-08-471-058-19	Sequence 19, Appl
30	68	6.8	280	3	US-08-471-057-19	Sequence 19, Appl
31	68	6.8	280	4	US-09-234-186-3	Sequence 3, Appl
32	68	6.8	280	5	PCT-US93-05651-3	Sequence 3, Appl
33	68	6.8	481	3	US-08-959-381A-1	Sequence 1, Appl
34	68	6.8	542	3	US-08-959-381A-2	Sequence 1, Appl
35	68	6.8	620	1	US-08-325-071-65	Sequence 65, Appl
36	68	6.8	620	4	US-08-461-004A-65	Sequence 65, Appl
37	68	6.8	650	1	US-08-325-071-67	Sequence 67, Appl
38	68	6.8	650	4	US-08-461-004A-67	Sequence 67, Appl
39	68	6.8	3079	5	PCT-US94-00198-4	Sequence 4, Appl
40	67.5	6.7	613	1	US-08-465-667A-2	Sequence 2, Appl
41	67.5	6.7	613	4	US-09-030-970-2	Sequence 2, Appl
42	67.5	6.7	613	5	PCT-US94-11843-2	Sequence 2, Appl
43	67	6.7	257	3	US-08-981-256A-12	Sequence 12, Appl
44	66.5	6.6	1023	1	US-08-198-446B-4	Sequence 4, Appl
45	66.5	6.6	1023	2	US-08-870-693-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-08-576-626A-42  
Sequence 42, Application US/08576626A  
Patent No. 5998194  
GENERAL INFORMATION:  
APPLICANT: Summers, R.G.  
APPLICANT: Katz, L.  
APPLICANT: Donadio, S.  
APPLICANT: Staver, M.J.  
TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR  
TITLE OF INVENTION: BIOSYNTHESIS GENES  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/576,626A  
FILING DATE: 21-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Diane Casulo  
REGISTRATION NUMBER: P-40,943  
REFERENCE/DOCKET NUMBER: 5857.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (847) 938-3137  
TELEFAX: (847) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 286 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5998194e  
US-08-576-626A-42  
Query Match 8.4%; Score 84; DB 2; Length 286;  
Best Local Similarity 22.2%; Pred. No. 0.051;

```

1  APPLICANT: GRODEN      METHODS FOR DIAGNOSIS AND TREATMENT
2  TITLE OF INVENTION:  OF BLOOM'S SYNDROME
3  NUMBER OF SEQUENCES:  78
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
6  STREET:  90 PARK AVENUE
7  CITY:  NEW YORK
8  STATE:  NEW YORK
9  COUNTRY:  U.S.A.
10 ZIP:  10016
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  3.5 INCH 1.44 MB STORAGE DISKETTE
14 COMPUTER:  IBM PC COMPATIBLE
15 OPERATING SYSTEM:  MS-DOS
16 SOFTWARE:  ASCII
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/559,303B
20 FILING DATE:  NOVEMBER 15, 1995
21 ATTORNEY/AGENT INFORMATION:
22 NAME:  ELIZABETH A. BOGOSIAN
23 REGISTRATION NUMBER:  39,911
24 REFERENCE/DOCKET NUMBER:  63475/65
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE:  (212) 697-5995
27 TELEFAX:  (212) 286-0854 or 286-0082
28 TELEX:  TWX 710-581-4766
29 INFORMATION FOR SEQ ID NO:  78:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH:  1417
32 TYPE:  AMINO ACID
33 STRANDEDNESS:  SINGLE
34 TOPOLOGY:  LINEAR
35
36 MOLECULE TYPE:
37 DESCRIPTION:  OTHER NUCLEIC ACID
38
39 HYPOTHETICAL:  YES
40 ANTI-SENSE:  NO
41 FEATURE:
42 NAME/KEY:
43 LOCATION:
44 IDENTIFICATION METHOD:
45 OTHER INFORMATION:
46
47 US-08-559-303B-78
48
49
50 Query Match      7.7%; Score 77; DB 2; Length 1417;
51 Best Local Similarity 31.8%; Pred. No. 4;
52 Matches 34; Conservative 21; Mismatches 32; Indels 20; Gaps 8;
53
54 QY      31 DFDY-WDVVPEPNNEVFEEST---CONVLKMLENCLSKSKORFKGSKVLY--PEKL 84
55      | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
56 Db      293 DDDYDTDPVPPSP--EELISASSSSSKLSTLKDIDT--SDRKEDVLSKSLKSPKPM 348
57      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 QY      85 TORIADVLRLSSTE----PCGLRGCVMHVNLIEINVCCKKIDRIYCD 127
59      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 Db      349 SM--QELNPTSTDCDARQLSLIQOOLIHV--MEHICKLIDTIPDD 389
61      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62
63 RESULT 4
64 US-08-781-891-78
65 ; Sequence 78, Application US/08781891
66 ; Patent No. 6090620
67 ; GENERAL INFORMATION:
68 ; APPLICANT:  Fu, Ying-Hui
69 ; APPLICANT:  Yu, Chang-Fn
70 ; APPLICANT:  Oshima, Junko
71 ; APPLICANT:  Mulligan, John T.
72 ; APPLICANT:  Schellenberg, Gerald D.
73 ; TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
74 ; TITLE OF INVENTION:  WERNER'S SYNDROME
75 ; NUMBER OF SEQUENCES:  209
76 ; CORRESPONDENCE ADDRESS:
77 ; ADDRESSEE:  SEED and BERRY LLP
78 ; STREET:  6300 Columbia Center, 701 Fifth Avenue
79

```

CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052,419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 682-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1417 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-78

Query Match 7.7%; Score 77; DB 3; Length 1417;  
Best Local Similarity 31.8%; Pred. No. 4;  
Matches 34; Conservative 21; Mismatches 32; Indels 20; Gaps 8;

Qy 31 DEDY-WDYVPEPPNINEVFEEST---CONLYKMLENCLSKSKOTKLGCSKYL--PEKL 84  
Db 293 DDDYDTPDVPSP--EETISASSSSKCLSTLKDLDT--SDRKEDVLSTSKDLSPKPM 348  
85 TORIADYLRISTE---PCGLRGCVMHVNEIENVCKKIDRIYCD 127  
Db 349 SM---QELNPETSTDCDARQISLQOQLIHV---MEHICKLIDITPDD 389

RESULT 5  
US-09-175-828-78  
Sequence 78, Application US/09175828  
Patent No. 6221643  
GENERAL INFORMATION:  
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA  
APPLICANT: GRODEN  
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT  
OF BLOOM'S SYNDROME  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRES:  
ADDRESSEE: AMSTER, ROTHSCHILD & EBBENSTEIN  
STREET: 90 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/175, 828  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/559, 303  
FILING DATE: NOVEMBER 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 63475/65

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1417  
TYPE: AMINO ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: OTHER NUCLEIC ACID  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
US-09-175-828-78

Query Match 7.7%; Score 77; DB 4; Length 1417;  
Best Local Similarity 31.8%; Pred. No. 4;  
Matches 34; Conservative 21; Mismatches 32; Indels 20; Gaps 8;

Qy 31 DEDY-WDYVPEPPNINEVFEEST---CONLYKMLENCLSKSKOTKLGCSKYL--PEKL 84  
Db 293 DDDYDTPDVPSP--EETISASSSSKCLSTLKDLDT--SDRKEDVLSTSKDLSPKPM 348  
85 TORIADYLRISTE---PCGLRGCVMHVNEIENVCKKIDRIYCD 127  
Db 349 SM---QELNPETSTDCDARQISLQOQLIHV---MEHICKLIDITPDD 389

RESULT 6  
US-08-325-071-59  
Sequence 59, Application US/08325071  
Patent No. 5587311  
GENERAL INFORMATION:  
APPLICANT: COBON, Stewart Gary  
APPLICANT: MOORE, Joanna Terry  
APPLICANT: JOHNSON, Law Anthony York  
APPLICANT: WILLADSEN, Peter  
APPLICANT: KEMP, David Harold  
APPLICANT: SRISKANTHA, Alagacone  
APPLICANT: RIDING, George Allred  
TITLE OF INVENTION: DNA Encoding A Cell Membrane  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRES:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/325, 071  
FILING DATE: 14-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/062,109  
FILING DATE: 17-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/926,368  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/242,196  
FILING DATE: 06-JUL-1988

1 APPLICATION NUMBER: US/08/461,004A  
2 FILING DATE: 04-JUN-1995  
3 PRIOR APPLICATION DATA:  
4 APPLICATION NUMBER: US 08/325,071  
5 FILING DATE: 19-OCT-1994  
6 PRIOR APPLICATION DATA:  
7 APPLICATION NUMBER: US 08/062,109  
8 FILING DATE: 17-MAY-1993  
9 PRIOR APPLICATION DATA:  
10 APPLICATION NUMBER: US 07/926,368  
11 FILING DATE: 07-AUG-1992  
12 PRIOR APPLICATION DATA:  
13 APPLICATION NUMBER: 07/242,196  
14 FILING DATE: 06-JUL-1988  
15 PRIOR APPLICATION DATA:  
16 APPLICATION NUMBER: PCT/AU87/00401  
17 FILING DATE: 27-NOV-1987  
18 PRIOR APPLICATION DATA:  
19 APPLICATION NUMBER: AU PI4912  
20 FILING DATE: 16-OCT-1987  
21 PRIOR APPLICATION DATA:  
22 APPLICATION NUMBER: AU PI2570  
23 FILING DATE: 19-JUN-1987  
24 PRIOR APPLICATION DATA:  
25 APPLICATION NUMBER: AU PH9196  
26 FILING DATE: 27-NOV-1986  
27 ATTORNEY/AGENT INFORMATION:  
28

```

:      REGISTRATION NUMBER: 29,768
:      REFERENCE/DOCKET NUMBER: 60042/152
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: 202 672 5300
:      TELEFAX: 202 672 5399
:      TELEX: 904136
:      INFORMATION FOR SEO ID NO: 59:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 650 amino acids
:      TYPE: amino acid
:      TOPOLOGY: linear
:      MOLECULE TYPE: protein
:
US-08-461-004A-59

```

[illegible]

Query Match	7.28;	Score 72.5;	DB 4;	Length 284
Best Local Similarity	23.88;	Pred. No. 1.3;		

```

Query Match      7.2%   Score 72; DB 4; Length 756;
Best Local Similarity 24.5%; Pred. No. 6.4;
Matches 49; Conservative 22; Mismatches 71; Indels 58; Gaps 10.
          12 PASSTFELDCGHPHSLSDPFYWDVY-----PEP--NUNEV-----47
          |         |||       |||         |||       :||
Db    401 PEDISELL---HSITLTAHLHT-GDTVIQGSATSLRAPAPDSLTACRQYGRETLAYL 455
          48 --IIEESTQQLN-VKALENCISK-----SKOTKLIG-----CSKYLVPKR 83

```

Storage  
Filed 05-27-1988

DB 456 SSLEEGETVENADVTALRNCLSRVKTGLGELLBRGLDIOKEELGDLVDKEMATSAIEA 515  
OY 84 LFORADVLRLSTEBCEGRGCVMHVNLIEIENWCKKIDRIYDSSVPPFEELTVFKOE 143  
DB 516 ATRIEE-----LISKRAGDTGKLEVENILISCTSLMOAT-KVLVASKDLOKEIVES 570  
OY 144 NCSWTSFRDFFFSRGRSSG 163  
DB 571 GRGSAPKPEFYAKNSRWTEG 590

## RESULT 11

US-08-775-009-34  
; Sequence 34, Application US/08775009  
; Patent No. 5935783  
; GENERAL INFORMATION:  
; APPLICANT: Gong, Wellong  
; APPLICANT: Emanuel, Beverly S.  
; APPLICANT: Budarf, Marcia L.  
; APPLICANT: Roe, Bruce  
; TITLE OF INVENTION: No. 5935783el Genes Mapping in the Digeorge and  
; TITLE OF INVENTION: Vellocardiofacial Syndrome Minimal Critical Region  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
; ADDRESSEE: No. 5935783rls, LLP  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/775,009  
; FILING DATE: 27-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trujillo, Doreen Yalko  
; REGISTRATION NUMBER: 35,719  
; REFERENCE/DOCKET NUMBER: CH-0681  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ. ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 871 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-775-009-34

Query Match 7.2%; Score 72; DB 2; Length 871;  
Best Local Similarity 27.2%; Pred. No. 7.9;

Matches 41; Conservative 24; Mismatches 52; Indels 34; Gaps 9;

OY 10 KNASISSEL-----LDCGYHESLSDPDYDYV-----PEPNLNEVFEESTCO-NL 57  
DB 107 EEPATHNALAKIYIDSNNSSECFLEKRENAVYDSSVGRYCKEKRDPHLACVAYERGQCDLEL 166  
OY 58 VKML-ENCLSKSKOTKIGCKSVLVP-----KLPQRIADVLRLSTBPCG 102  
DB 167 IKCNENSLFKSARLYVCK--DPELMAVLEETNPSRQLIDYVOTALS-ETRDPEE 223  
OY 103 LRGCV--MHVNLIEIENWCKKIDRIYDSSV 130  
DB 224 ISVTYKAFMTADLPNE-LIELLEKIYIDNSV 253

RESULT 12  
US-09-180-852-2  
; Sequence 2, Application US/09180852  
; Patent No. 6287834  
; GENERAL INFORMATION:  
; APPLICANT: BELANGER, Alain  
; APPLICANT: HUM, Dean W.  
; APPLICANT: BEAULIEU, Martin  
; APPLICANT: LEVESQUE, Eric  
; TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE  
; TITLE OF INVENTION: DIPHOSPHO-GLUCURONOSYLTRANSFERASE  
; FILE REFERENCE: 1259-449  
; CURRENT APPLICATION NUMBER: US/09/180,852  
; CURRENT FILING DATE: 1999-02-08  
; EARLIER APPLICATION NUMBER: PCT/CA97/00328  
; EARLIER FILING DATE: 1997-05-16  
; EARLIER APPLICATION NUMBER: US 08/649,319  
; EARLIER FILING DATE: 1996-05-17  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ. ID NO 2  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-180-852-2

Query Match 7.0%; Score 70.5; DB 4; Length 530;  
Best Local Similarity 28.6%; Pred. No. 5.7;

Matches 34; Conservative 12; Mismatches 42; Indels 31; Gaps 6;

OY 1 MVATGSLSKNPNASISLELDCGYHPSL-----LSD-----FDYWDYVPE-----PNL 44  
DB 54 IYLTSSASILVNAKSSSAILEVYPTSLTKNDLEDFPMKMFDBWTVSISKRTWYSFSQL 113  
OY 45 NEVFEESTCONLYKMLENC-LSKSKOTKIGCKSVLPEKLTQRIADVLRLSTBPCG 102  
DB 114 QELCWETSDYN--IKLCEADVLNKKMKRLQESKF-----DVLADAVNFCG 158

## RESULT 13

US-08-890-865A-10  
; Sequence 10, Application US/08890865A  
; Patent No. 6307019  
; GENERAL INFORMATION:  
; APPLICANT: Constantini, Franklin  
; APPLICANT: Zeng, Li  
; TITLE OF INVENTION: AXIN GENE AND USPS THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: US  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/890,865A  
; FILING DATE: 10-JUL-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 0575/54249  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)278-0400  
; TELEFAX: (212)391-0526  
; INFORMATION FOR SEQ. ID NO: 10:

Query Match 7.2%; Score 72; DB 2; Length 871;  
Best Local Similarity 27.2%; Pred. No. 7.9;  
Matches 41; Conservative 24; Mismatches 52; Indels 34; Gaps 9;

APPLICATION NUMBER: AU P12370  
FILING DATE: 19-TIN-1097

FILED DATE: 1/26/1993  
PRIOR APPLICATION DATA:



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 12:55:07 ; Search time 35 Seconds  
(without alignments)  
213.511 Million cell updates/sec

Title: US-09-881-636-2

Perfect score: 1002  
Sequence: 1 MVATGSLSSKNPASISELD.....FRVKKRLYSLIGTTVIEGS 193

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90.5	9.0	1056	RIC1_YEAST	P40395 Saccharomyces cerevisiae
2	85	8.5	1098	RP16_DROME	O94900 drosophila
3	80	8.0	1416	BLM_MOUSE	O88700 mus musculus
4	79	7.9	315	CCG3_HUMAN	O60359 homo sapien
5	79	7.9	549	CEB4_CAEEL	P30429 caenorhabditis
6	78.5	7.8	463	GLGA_AQUAE	O60935 aquilex aeo
7	78.5	7.8	964	ION2_MAIZE	P93648 zea mays (m
8	77.5	7.7	356	Y4ZB_RHISE	P55729 rhizobium s
9	77	7.7	1417	BLM_HUMAN	P54132 homo sapien
10	77	7.7	1828	MY5A_RAT	O94973 rattus norv
11	77	7.7	1829	MY5A_CHICK	O02440 gallus galli
12	77	7.7	1853	MY5A_MOUSE	O99104 mus musculu
13	77	7.7	1855	MY5A_HUMAN	O99411 homo sapien
14	76	7.6	1390	RPOB_MYCGA	P47715 mycoplasma
15	75.5	7.5	613	GP37_HUMAN	O13354 homo sapien
16	75	7.5	418	FTSA_BUCAP	O51928 buchnera ap
17	75	7.5	439	VP51_BRAPS	O911P7 bacterioph
18	74.5	7.4	323	CCG2_HUMAN	O97698 homo sapien
19	74.5	7.4	323	CCG2_MOUSE	O88602 mus musculu
20	74.5	7.4	676	TIM_DROHY	O44431 drosophila
21	74.5	7.4	4196	DYHC_SCHPO	O13290 schizosacch
22	73.5	7.3	366	FLHF_BACSU	O01960 bacillus su
23	73.5	7.3	441	VP40_HUMAN	O952K6 homo sapien
24	73	7.3	145	RI15_BORBU	O51450 borrelia bu
25	73	7.3	334	DCAM_BOVIN	P50243 bos taurus
26	73	7.3	334	DCAM_MESAU	P28918 mesocricetu
27	73	7.3	664	SL51_SHEEP	P53791 ovvis aries
28	72.5	7.2	610	VE1_HPV60	O80943 human papil
29	72	7.2	334	DCAM_HUMAN	P17707 human saplen
30	72	7.2	486	VNS3_RORHM	P35424 human rotav
31	72	7.2	530	UDB2_RAT	P08541 rattus norv
32	72	7.2	1640	CUH2_HUMAN	P53675 homo sapien
33	71.5	7.1	410	Y474_AQUAE	O66772 aquifex aeo

34	71.5	7.1	510	1	GARD_BACSU	P42240 bacillus su
35	71.5	7.1	566	1	TS13_MOUSE	O01755 mus musculu
36	71.5	7.1	1818	1	2294_HUMAN	O94822 homo sapien
37	71	7.1	179	1	RL5_BUCAP	P46178 buchnera ap
38	71	7.1	469	1	VL2_BPV1	P03109 bovine papil
39	71	7.1	814	1	STV_RICPR	O94906 rickettsia
40	71	7.1	1005	1	EAL2_HUMAN	O43491 homo sapien
41	70.5	7.0	333	1	DCAM_RAT	P17708 rattus norv
42	70.5	7.0	334	1	DCM1_MOUSE	P31154 mus musculu
43	70.5	7.0	334	1	DCM2_MOUSE	P82184 mus musculu
44	70.5	7.0	334	1	DCM2_MOUSE	P82185 mus sapien
45	70.5	7.0	530	1	UDBH_HUMAN	O75795 homo sapien

## ALIGNMENTS

## RESULT 1

ID	RIC1_YEAST	STANDARD:	PRT:	1056 AA.
AC	P40395;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	RIC1 protein.			
GN	RIC1 OR YLR039C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Mizuta K.;			
RL	Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Koetter P., Rose M., Ertan K.D.;			
RL	Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.			
CC	-I- FUNCTION: INVOLVED IN THE TRANSCRIPTION OF BOTH RIBOSOMAL PROTEIN			
CC	GENES AND RIBOSOMAL RNA.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: DA3895; BAA07866.1; -			
DR	EMBL: Z73211; CAA97567.1; -			
DR	TRANSFAC: T03560; -			
DR	SGD: S0004029; RIC1.			
RN	Transcription regulation.			
SQ	SEQUENCE 1056 AA; 121649 MW; 2368616B67AEAD31 CRC64;			

Query Match 9.0%; Score 90.5; DB 1; Length 1056;

Best Local Similarity 25.0%; Pred. No. 1.3;

Matches 42; Conservative 23; Mismatches 52; Indels 51; Gaps 8;

QY	24	HPESLSDPDYDYYVPEPNLNEVFEESTQNLVMLENCSKSKOTLGSKYLVPRK	83
DB	613	NDVDELSDVYVMEYVNP-----NTIYF-----SLHYNTLSRKLLKMKSKNHNASKR	660
QY	84	L-----TORIADVLRLSTPECGLRG-----VMHVNLEINVCCKLDR	123
DB	661	QPDALIKTAELILVDTQIVFDV--ISTVHPCGLNIIRKFFQYLIKINIPVLPKIKW	718
QY	124	ICYDSSVTFPELTVFKENCSSWTSFRD--FFESRGRFSSGRRRTIL	170
DB	719	II-----NMKEGLIF-----FADRKRIKLGKVDGGMQITLT	750

RESULT 2  
 RPL6\_DROME STANDARD: PRT: 1098 AA.  
 AC O9GON0: O9VXL6; (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Ran-binding protein 16.  
 GN RANBP16 OR CG9126.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Ephygryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NX NCBI\_TaxID=7227;  
 RP SEQUENCE FROM N.A.  
 RL Goerlich D., Hartman E.;  
 Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN  
 RC STRAIN-Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champé M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Chertis J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry K.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Mostreli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -I- FUNCTION: May function as a nuclear transport receptor (By  
 similarity).  
 CC -I- SUBUNIT: Binds to nucleoporins and the GTP-bound form of Ran (By  
 similarity).  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic, nuclear and nuclear pore  
 complex (By similarity).  
 CC -I- SIMILARITY: BELONGS TO THE EXPORTIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF222744; AAG44254.1; -.  
 DR EMBL: AF003500; AAF48542.1; ALT\_SEQ.  
 DR FLYBASE: FBgn0030709; RANBP16.  
 KW Nuclear protein; Transport; Protein transport.  
 SQ SEQUENCE 1098 AA; 125532 MW; C8A385CB3FPAE5 CRC64;  
 Query Match 8.5%; Score 85; DB 1; Length 1098;  
 Best Local Similarity 25.6%; Pred. No. 4.5;  
 Matches 30; Conservative 21; Mismatches 44; Indels 22; Gaps 4;  
 QY 36 DYVVPENLNEVFEESTCONLTKMLENCLSKSKOTRLGCSKVLVPEKLTQRIADYLR 95  
 DB 83 NYLATVPNLQHFV-----VQALVSL-----AKLRKGFWDGSKEMVQNLLEDYKRF 131  
 QY 96 SSTEPCGLRGCAVHVNLEIENCKKLDRIYCDSSVPTFELTLVFKOENCSWTSPFD 152  
 DB 132 -----LQGSVBHCITIGVQ-----ILSQLVCEMNSVEMDVQVFSKMRKIATSPFD 177  
 RESULT 3  
 BLM\_MOUSE STANDARD: PRT: 1416 AA.  
 AC O88700; O88198;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Bloom's syndrome protein homolog (EC 3.6.1.-) (bBLM).  
 GN BLM.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN  
 RP SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS.  
 RX MEDLINE=99054654; PubMed=9840919;  
 RA Bahr A., de Graeve F., Keding C., Chatten B.;  
 RT "Point mutations causing Bloom's syndrome abolish ATPase and DNA  
 RT helicase activities of the BLM protein.";  
 RL Oncogene 17:2565-2571(1998).  
 RN  
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
 RC STRAIN-BALB/C; TISSUE-Testis, Spermatoocyte, and Brain;  
 RX MEDLINE=98322127; PubMed=9655940;  
 RA Seki T., Wang W.-S., Okumura N., Seki M., Katada T., Enomoto T.;  
 RT "cDNA cloning of mouse BLM gene, the homologue to human Bloom's  
 RT syndrome gene, which is highly expressed in the testis at the mRNA  
 RT level.";  
 RL Biochim. Biophys. Acta 1398:377-381(1998).  
 CC -I- FUNCTION: PARTICIPATES IN DNA REPLICATION AND MAY PARTICIPATE IN  
 CC REPAIR. EXHIBITS A MAGNESIUM-DEPENDENT ATP-DEPENDENT DNA-HELICASE  
 CC ACTIVITY THAT UNWINDS SINGLE- AND DOUBLE-STRANDED DNA IN A 3'-5'  
 CC DIRECTION.  
 CC -I- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS 12-14 DAYS AFTER  
 CC BIRTH (CORRESPONDING TO THE PACHYTENE PHASE) AND AT MUCH LOWER  
 CC LEVELS IN BRAIN, HEART, KIDNEY, LUNG, THYMUS, KIDNEY AND SPLEEN.  
 CC -I- SIMILARITY: BELONGS TO THE HELICASE FAMILY. REG SUBFAMILY.  
 CC -I- SIMILARITY: CONTAINS 1 HDRC DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: Z98263; CAB10933.1; -.  
 DR EMBL: AB008674; BAA32001.1; -.  
 DR MGD: MGI:1328362; BLM.

```

DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002464; DEAD_ATP_helicase.
DR InterPro: IPR002121; HRDC.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD_1.
DR Pfam: PF00271; helicase_C_1.
DR Pfam: PF00570; HRDC_1.
DR SMART: SM00487; DEXDC_1.
DR SMART: SM00490; HELICC_1.
DR SMART: SM00341; HRDC_1.
DR PROSITE: PS00690; DEAD_ATP_HELICASE_1.
KM Hydrolyase; Helicase; ATP-binding; DNA-binding; Nuclear protein;
KW DNA replication.
FT DOMAIN 219 231 POLY-GLU.
FT DOMAIN 564 574 POLY-ASP.
FT DOMAIN 1312 1317 POLY-GLU.
FT NP_BIND 697 704 ATP (BY SIMILARITY).
FT SITE 803 806 DEAD BOX.
FT DOMAIN 1217 1297 HDRC.
FT DOMAIN 1333 1348 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MUTAGEN 680 680 Q->P: REDUCED ATPASE AND HELICASE
FT MUTAGEN 703 703 K->A: REDUCED ATPASE AND HELICASE
FT MUTAGEN 849 849 I->T: REDUCED ATPASE AND HELICASE
FT MUTAGEN 1063 1063 ACTIVITIES.
FT CONFLICT 131 131 C->S: REDUCED ATPASE AND HELICASE
FT CONFLICT 229 229 ACTIVITIES.
FT CONFLICT 535 535 L->P (IN REF. 2).
FT CONFLICT 546 547 E->E (IN REF. 2).
FT CONFLICT 574 574 V->M (IN REF. 2).
FT CONFLICT 591 591 WN->RT (IN REF. 2).
FT CONFLICT 591 591 MISSING (IN REF. 2).
FT CONFLICT 621 621 T->A (IN REF. 2).
FT CONFLICT 1295 1295 T->N (IN REF. 2).
FT CONFLICT 1416 1416 V->L (IN REF. 2).
SQ SEQUENCE 1416 AA; 158365 MW; 447C8110A775DD42 CRC64;

Query Match 8.0%; Score 80; DB 1; Length 1416;
Best Local Similarity 24.5%; Pred. No. 18;
Matches 38; Conservative 32; Mismatches 43; Indels 42; Gaps 11;

QY 30 SDPDYWDYVPEPNIENYFEEST---CONLVKMLENCLSKSKOTKLGSKVLY--PEKL 84
DB 299 NDYDI-DEVPSP--ELITSAASSLKSSMKLDD--SKEKGIISTEELSKPEEM 353
QY 85 TORIADYVLRISTEPCG-----LRGCVHVNLEIENVCKLRIVCDSSVVPFETLV 139
DB 354 TTHKSD---AGTSKDCDAQIIRIQOOLIV--MEHICKLV-----TVPTELAL 399
QY 140 FKQENCSWTSFRDFFFSRGRSSGFRRLTLLSSGF 174
DB 400 -----NCGTELLQ-----QRNIRKRLADAGP 421

RESULT 4
CCG3_HUMAN STANDARD: PRT; 315 AA.
AC 060359;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-dependent calcium channel gamma-3 subunit (Neuronal voltage-
DE gated calcium channel gamma-3 subunit).
GN CACNG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RX MEDLINE=99236653; PubMed=10221464;

```

```

RA Black J.L. III, Lennon V.A.;
RT "Identification and cloning of putative human neuronal voltage-gated
RT calcium channel gamma-2 and gamma-3 subunits: neurologic
RT implications.";
RL Mayo Clin. Proc. 74:357-361(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
RL Genomics 60:295-308(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain cortex;
RA Xia J.-H., Zhang H.-L., Tang X.-X., Yu K.-P., Pan Q., Dai H.-P.;
RT "Cloning of human calcium channel gamma-3 subunit.";
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: THOUGHT TO STABILIZE THE CALCIUM CHANNEL IN AN
CC -I- INACTIVATED (CLOSED) STATE.
CC -I- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FIVE SUBUNITS:
CC -I- ALPHA-1, ALPHA-2/DELTA, BETA AND GAMMA.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- DISEASE: CANDIDATE GENE FOR A FAMILIAL INFANTILE CONVULSIVE
CC -I- DISORDER WITH PAROXYSMAL CHOREOATHETOSIS.
CC -I- SIMILARITY: BELONGS TO THE PMR-22 / EMP / MP20 FAMILY. CACNG
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF100346; AAD22739.1; -
CC EMBL: AC004125; AAC15246.1; -
CC EMBL: AF134640; AAF42975.1; -
CC MIM: 606403; -
DR InterPro: IPR004031; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
SQ SEQUENCE 315 AA; 35548 MW; 5DCFB9FBA4B5A2F4 CRC64;

Query Match 7.9%; Score 79; DB 1; Length 315;
Best Local Similarity 23.4%; Pred. No. 3.6;
Matches 41; Conservative 16; Mismatches 50; Indels 68; Gaps 8;

QY 33 DYWDYVPEPNIENYFEESTCONLVKMLENCLSKSKOTKLGSKVLYPEKLTORIADYV 92
DB 31 DYV-----LYSRGYC-----RTKSTYS-----RIYDDSSVVPFF 57
QY 93 LRLST-EPGGLRGCVHVNLEIENVCKLKD-----RIVDDSSVVPFF 134
DB 58 MHSLSMTRCTLEGA-----FRGVCKIDHFPEDADYEQDTAEYLRAVYASSVFPFL 110
QY 135 ELTVFKQENCSWTSFRDFFFSRGRSSGFRRLTLLSSGFRLLVKKYLSLIGTTV 189
DB 111 SVTLFLFGFLCYAAS--EFHRS-----RHNVILSGIFVSVAGLSNIIIGIIV 155

RESULT 5
CED4_CAEBL STANDARD: PRT; 549 AA.
ID CED4_CAEBL

```

```
AC P03429; 1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell death protein 4.
GN CED-4 OR C35D10.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Pelodermineae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93161939; PubMed=1286611;
RA Yuan J., Horvitz H.R.;
RT "The Caenorhabditis elegans cell death gene ced-4 encodes a novel
RT protein and is expressed during the period of extensive programmed
RT cell death."
RL Development 116:309-320(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton L.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: EXPRESSED DURING THE PERIOD OF EXTENSIVE PROGRAMMED
CC CELL DEATH. CED-4, ALONG WITH CED-3, SEEMS TO FUNCTION WITHIN
CC DYING CELLS TO CAUSE CELL DEATH.
CC -1- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING EMBRYOGENESIS AND IS
CC ALSO DETECTED AT LATER STAGES.
CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; X69016; CAA48781.1; -.
DR EMBL; U21324; AAA62564.1; -.
DR WormPep; C35D10.9; CE01203.
DR InterPro; IPR001315; CARD.
DR InterPro; IPR002182; NB-ARC.
DR Pfam; PF00619; CARD. 1
DR Pfam; PF00931; NB-ARC; 1.
DR SMART; SM00114; CARD; 1.
DR PROSITE; PS0209; CARD; 1.
KW Apoptosis; ATP-binding.
FT DOMAIN 1 91 CARD.
FT NP_BIND 159 166 ATP (POTENTIAL).
FT FT 549 AA; 62878 MW; DBA7969DA50AF8 CRC64;
SO SEQUENCE

Query Match 7.9%; Score 79; DB 1; Length 549;
Best Local Similarity 24.7%; Pred. No. 7;
Matches 45; Conservative 23; Mismatches 46; Indels 68; Gaps 11.
```

RESULT 6			
ID	GIGA_AQUAE	STANDARD:	PRT: 463 AA.
AC	069935;		
DT	01-MAR-2002 (Rel. 41, Created)		
DT	01-MAR-2002 (Rel. 41, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Glycogen synthase (EC 2.4.1.21) (Starch [bacterial glycogen] synthase)		
DE	GIGA OR AQ_721.		
OS	Aquifex aeolicus.		
OC	Bacteria; Aquificales; Aquificaceae; Aquifex.		
OX	NCBI_taxid=63363;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VF5;		
RX	MEDLINE=9819666; PubMed=9537320;		
RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;		
RT	"The complete genome of the hyperthermophilic bacterium Aquifex aeolicus";		
RL	Nature 352:353-358(1998).		
CC	-1- FUNCTION: Synthesizes alpha-1,4-glucan chains using ADP-glucose.		
CC	-1- CATALYTIC ACTIVITY: ADP-glucose + ((1,4)-alpha-D-glucosyl)(N) =		
CC	ADP + ((1,4)-alpha-D-glucosyl)(N+1).		
CC	-1- PATHWAY: Glycogen biosynthesis; second step.		
CC	-1- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; AE000704; AAC06894.1; -		
DR	InterPro; IPR001296; Glycos_transf_1.		
DR	Pfam; PF00534; Glycos_transf_1; 1.		
KW	Glycogen biosynthetases; Transferase; Glycosyltransferase;		
KW	Complete proteome.		
FT	BINDING 15 ADP-GLUCOSE (BY SIMILARITY).		
SQ	SEQUENCE 463 AA; 53457 MW; DB96172DD03DB27D CXC64;		
Query Match 7.8%; Score 78.5; DB 1; Length 463;			
Best local similarity 19.6%; Pred. No. 6.4;			
Matches 37; Conservative 37; Mismatches 78; Indels 37; Gaps 5;			
QY	1 MYATGSLSSKKRPASISLSEL-LDCGHPHSLSLDPY-----WDYVVPENLNVETFE 50		
Db	198 IYFSDLIITVVPYAKELIQTQDEYGGELGVLAKKTSYKRLGLINGIDYEWNNEKKYIYQ 257		
QY	51 ESTCQNTKVMLENCISLSSKQTKLGCSKYLV--EKLTORIADQVLRSLSTPECGILGCGVM 108		
Db	258 NYSLRNYSKRRKRNKEFLSKELGIEAKRLISINFTHQKGVEL----- 301		
QY	109 HVNLEIENVCKKLDRIYCDSSVPTFELLTVFKQENCWMTSFRDFEGRSGCFRRTL 168		
Db	302 -----ILNCAEEMSKLTANFVFLGTGEYENAFLDYSKIYKNEKVP-----AEFNEGFARRL 352		
QY	169 ILSSGFRLV 177		
Db	353 YASDFEILM 361		

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lon protease homolog 2, mitochondrial precursor (EC 3.4.21.-).
GN LON2.
OS zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
ON NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_873;
RA Rapp W.D., Barakat S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE LON
CC FAMILY OF ATP-DEPENDENT PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U85495; AAC50021.1; -
DR MEROPS: S16.002; -
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003959; AAA_subfam.
DR InterPro: IPR003111; LON.
DR InterPro: IPR001984; Lon_endopep.
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF02190; LON; 1.
DR PRINTS: PR00830; ENDOLAPTASE.
DR SMART: SM00382; AAA; 1.
DR SMART: SM00464; LON; 1.
DR PROSITE: PS01046; LON_SER; 1.
DR Hydroxylase; Serine protease; ATP-binding; Multigene family;
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 964 LON PROTEASE HOMOLOG 2.
FT NP_BIND 455 462 ATP (POTENTIAL).
FT ACT_SITE 863 863 BY SIMILARITY.
SQ SEQUENCE 964 AA; 105659 MW; D78AC6B0F8A6D9E CRC64;

Query Match 7.8%; Score 78.5; DB 1; Length 964;
Best Local Similarity 27.6%; Pred. No. 15;
Matches 34; Conservative 28; Mismatches 42; Indels 19; Gaps 8;

QY 5 GSLSSKNPAS-ISELLDCGYHPESELISDFDYWDYVVPENLNEVFEESTC-QNLYKMLE 62
DB 528 GKHSNGDPASALVELD-----PEQNVNPLDH--YLDVPLDLSKYL--VCYANVIEKIP 578
QY 63 NCLSKSKOTKLGCSKVLVPEKLTQRIADVLRLSTEGCGKGVHVN-----LEIENV 117
DB 579 NPL-LDRMEITAIAGYIDDKW--HIARDYLEKNTROACGIKPEQYEVTFALLALIENV 635
QY 118 CKK 120
DB 636 CRE 638

RESULT 8
Y42B_RHISN STANDARD: PRT: 356 AA.
AC P55729;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative transposase Y42B.
GN Y42B.
OS Rhizobium sp. (strain NGR234).
OG plasmid sym PNGR234d.

```

```

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
ON NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: BELONGS TO THE TRANSPOSAE FAMILY 11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000108; AB91960.1; -
DR InterPro: IPR002559; Transposase_11.
DR Pfam: PF01609; Transposase_11; 1.
KW Hypothetical protein; Transposable element; Transposition;
KW DNA-binding; DNA recombination; Plasmid
SQ SEQUENCE 356 AA; 39775 MW; 9F350ABB7E691635 CRC64;

```

```

Query Match 7.7%; Score 77.5; DB 1; Length 356;
Best Local Similarity 25.2%; Pred. No. 5.7;
Matches 28; Conservative 22; Mismatches 34; Indels 27; Gaps 6;

```

```

QY 82 EKLQRIADVLRLSTEP-----CG-----LGCYHVNLEIENVCKKL---DRI 124
DB 76 DQIQRREGAMRLIDSTPIPLKLGMAKSNGRIGMKMHVYDDSDCPRLDITDAN 135
QY 125 VCDSSVPPPEL-----TLVFKQENCN---WTSFRD---PFSSGRSSGFR 165
DB 136 VADAQIGRTIAIESGATYIFDKGYCHYGMWTAIAEKAFVTRPKSNMGLK 186

```

```

RESULT 9
BLM_HUMAN STANDARD: PRT: 1417 AA.
ID BLM_HUMAN
AC P54132;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Bloom's syndrome protein (EC 3.6.1.-) (RecQ protein-like 3) (DNA
DE helicase, RecQ-like, type 2).
GN BLM OR RECQ3 OR RECQ2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS BS ARG-672; ILE-843 AND SER-1055.
RX MEDLINE=96069866; PubMed=7585968;
RA Ellis N.A., Groden J., Ye T.-Z., Straughen J., Lennon D.J., Ciccoci S.,
RA Prolycheva M., German J.;
RT "The Bloom's syndrome gene product is homologous to RecQ helicases.";
RL Cell 83:655-666(1995).
CC [2]
CC SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=B-cell;
RX MEDLINE=98049515; PubMed=9388193;
RA Karrow J.K., Chakraverly R.K., Hickson I.D.;
RT "The Bloom's syndrome gene product is a 3'-5' DNA helicase.";
RL J. Biol. Chem. 272:30611-30614(1997).
RN [3]
RP NUCLEAR LOCALIZATION SIGNAL.
RX MEDLINE=98049834; PubMed=9388480;
RA Kaneko H., Orli K.O., Matsui E., Shinozawa N., Fukao T., Matsumoto T.,
RA Shimamoto A., Furuichi Y., Hayakawa S., Kasahara K., Kondo N.;

```

Query Match	Score 77;	DB 1;	Length 1417;
Best Local Similarity	31.88;	Pred No. 34;	
RT			"BIM (the causative gene of Bloom syndrome) protein translocation into the nucleus by a nuclear localization signal".;
RL			Biochem. Biophys. Res. Commun. 240:348-353(1997).
RN			[4]
RA			VARIANT BS PHE-1036.
RX			MEDLINE-97449163; PubMed-9285778;
RA			Foucault F., Vauzy C., Barakat A., Thibout D., Planchon P., Jaulin C., Pratz F., Amor-Gueret M.;
RT			"Characterization of a new BIM mutation associated with a topoisomerase II alpha defect in a patient with Bloom's syndrome.";
RL			Hum. Mol. Genet. 6:1427-1434(1997).
CC			-1- FUNCTION: PARTICIPATES IN DNA REPLICATION AND MAY PARTICIPATE IN REPAIR. EXHIBITS A MAGNESIUM-DEPENDENT ATP-DEPENDENT DNA-HELICASE ACTIVITY THAT UNWINDS SINGLE- AND DOUBLE-STRANDED DNA IN A 3'-5' DIRECTION.
CC			-1- SUBCELLULAR LOCATION: Nuclear.
CC			-1- DISEASE: DEFECTS IN BIM ARE THE CAUSE OF BLOOM SYNDROME (BS), AN AUTOSOMAL RECESSIVE DISORDER CHARACTERIZED BY PROPORTIONATE PRE- AND POSTNATAL GROWTH DEFICIENCY, SUN-SENSITIVE, TELANGIECTATIC, HYPO- AND HYPERPIGMENTED SKIN; PREDISPOSITION TO MALIGNANCY, AND CHROMOSOMAL INSTABILITY.
CC			-1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECO SUBFAMILY.
CC			-1- SIMILARITY: CONTAINS 1 HDRC DOMAIN.
CC			-1- DATABASE: NAME=BIMbase; NOTE=BIM mutation db; WWW="http://www.uta.fi/lmc/bioinfo/BIMbase/"
CC			-----
CC			This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).
CC			-----
DR			EMBL; U39817; AAA87850.1; -
DR			MIM; 604610; -
DR			MIM; 210900; -
DR			InterPro: IPR001410; DEAD.
DR			InterPro: IPR002464; DEAD_ATP_helicase.
DR			InterPro: IPR002121; HRDC.
DR			InterPro: IPR001650; Helicase_C.
DR			Pfam; PF00270; DEAD; 1.
DR			Pfam; PF00271; helicase_C; 1.
DR			Pfam; PF00570; HRDC; 1.
DR			SMART; SM00487; DEXDC; 1.
DR			SMART; SM00490; HELIC; 1.
DR			SMART; SM00341; HRDC; 1.
DR			PROSITE: PS00690; DEAD_ATP_HELICASE; 1.
KW			Hydrolase; Helicase; ATP-binding; DNA-binding; Nuclear protein;
KW			DNA replication; Disease mutation.
FT			DOMAIN 292 299 POLY-ASP.
FT			DOMAIN 310 316 POLY-SER.
FT			DOMAIN 557 566 POLY-ASP.
FT			NP_BIND 689 696 ATP (BY SIMILARITY).
FT			STATE 795 798 DEAD BOX.
FT			DOMAIN 1212 1292 HRC.
FT			DOMAIN 1334 1349 HRC.
FT			VARIANT 672 672
FT			VARIANT 843 843
FT			VARIANT 891 891
FT			VARIANT 901 901
FT			VARIANT 1036 1036
FT			VARIANT 1055 1055
FT			SEQUENCE 1417 AA; 158999 MW; 423DF5F381194E11 CRC64;

```

Matches 34, Conservative 21, Mismatches 32, Indels 20, Gaps 8,
OY 31 PDDY-WDYVVEPPNINEXVIFEEEST---CONLYKMLENLSKSKOTKCLCCSKYLV--PEKL 84
Db 293 DDDYDVTDPVPSP--EEIISASSSSSKCLSTLKLDLT--SQRKEVLSSTSDLLSKPEKM 348
OY 85 TORIADVLRLSTSE----PCGLRCACVHMVNIENWCKKIDRIYCD 127
Db 349 SM---QELNPETSTDCDARQISLQOOLHV---MEHICKLIDTIPDD 389

RESULT 10
MY5A_RAT MY5A_RAT STANDARD: PTR: 1828 AA.
AC 090YF3:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Va (Myosin 5a) (Dilute myosin heavy chain, non-muscle).
GN MYO5A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
XX [1]
RN SEQUENCE FROM N.A.
RP Futaki S., Murata Y., Hayashi Y.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROGRESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC MAY BE INVOLVED IN MELANOSOME TRANSPORT. OR ALTERNATIVELY, IT MAY
CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC FORMATION (BY SIMILARITY).
CC -1- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 IO DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AB035736; BAA8350.1; -.
CC DR HSSP; P08799; ILVK.
CC DR InterPro; IPR002710; DIL.
CC DR InterPro; IPR000048; IO.
CC DR InterPro; IPR001609; myosin_head.
CC DR Pfam; PF01843; DIL.1.
CC DR Pfam; PF00612; IO.6.
CC DR Pfam; PF00063; myosin_head.1.
CC DR PRINTS; PR00193; MYOSINHEAVY.
CC DR ProDom; PD000355; myosin_head.1.
CC DR ProDom; PD003376; DIL.1.
CC DR SMART; SM00015; IO.6.
CC DR SMART; SM00242; MYSC.1.
CC DR PROSITE; PS50096; IO.6.
CC DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
CC Coiled coil; Phosphorylation.
CC KW Coiled coil; 1.
CC FT DOMAIN 1 765 MYOSIN HEAD-LIKE.
CC FT 766 788 IO 1.
CC FT 789 813 IO 2.
CC FT 814 836 IO 3.
CC FT 837 861 IO 4.
CC FT 862 884 IO 5.
CC FT 885 914 IO 6.
CC FT 914 1239 COILED COIL (POTENTIAL).
CC FT 1314 1418 COILED COIL (POTENTIAL).
CC FT 1660 1765 DILUTE.

```

```

FT NP_BIND 163 170 ATP (POTENTIAL).
FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1733 1733 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 1828 AA; 211762 MW; 5B3DE1C89AE36123 CRC64;

Query Match
Best Local Similarity 21.8%; Pred. No. 46;
Matches 42; Conservative 33; Mismatches 60; Indels 58; Gaps 9;

QY 26 ESLSDFPDVMDYVPEPNEVFEESTQNLVKMLNCLSKQTKLGCCKVLPKLT 85
D 1560 EHCLTNFPLAEYROYLSDLAIQIY-----QQLVRYLENTL-----QPMIYSGMIE 1604
QY 86 QRIADVLRSLSTPEPCGLRGCVHVNLE---IENVCKRLD---RIVCDSSVVPFEL-- 136
D 1605 HETIOGV---SGVKPTGLRKRTSSIADEGTYLDSILRLQNSFHSVWCQCHGMDP--ELIK 1659
QY 137 -----TLVFKQENCSW-----TSFRDFEFGFGFSGFRRT--- 167
D 1660 QVVKOMFYIYGAITLNNLLRLKDMCSKQMOIRYNVSOLEEMLKDKNLNMSGAKETLEP 1719
QY 168 LILSSGFRLVKK 180
D 1720 LIGAAQLLVKKK 1732

RESULT 11
MYSA_CHICK STANDARD; PRT; 1829 AA.
ID MYSA_CHICK
AC 002440:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Va (Myosin 5a) (Dilute myosin heavy chain, non-muscle) (Myosin
GN MYO5A.
OS Gallus gallus (chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID:9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain:
RX MEDLINE-93012002; PubMed-1383040;
RA Sanders G., Lichte B., Meyer H.E., Killmann M.W.;
RT "cDNA encoding the chicken ortholog of the mouse dilute gene product.
RT Sequence comparison reveals a myosin I subfamily with conserved C-
RT terminal domains."
RL FEBS Lett. 311:295-298(1992).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.
RC TISSUE-Brain:
RX MEDLINE-93107155; PubMed-1469047;
RA Espreffio E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
RA de Camilli P.V., Larson R.E., Mooseker M.S.;
RT "Primary structure and cellular localization of chicken brain
RT myosin-V (p190), an unconventional myosin with calmodulin light
RT chains."
RL J. Cell Biol. 119:1541-1557(1992).
CC -1- FUNCTION: PROGRESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC FORMATION (BY SIMILARITY).
CC -1- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES
CC (POTENTIAL).
CC -1- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X67251; CAA47673.1; -
CC EMBL: 211718; CA877782.1; -
CC HSPB, P08799; 1MMG.
CC InterPro: IPR002710; DIL.
CC InterPro: IPR000048; IQ.
CC Pfam: PF01843; DIL; 1.
CC Pfam: PF00612; IQ; 6.
CC PRINTS: PR00193; MYOSINHEAD.1.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 6.
CC SMART: SM00242; MYSC; 1.
CC PROSITE: PS50096; IQ; 6.
CC Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
CC Coiled coil; Phosphorylation.
CC MYOSIN HEAD-LIKE.
CC FT DOMAIN 1 766
CC FT DOMAIN 767 789
CC FT DOMAIN 790 814
CC FT DOMAIN 815 837
CC FT DOMAIN 838 862
CC FT DOMAIN 863 887
CC FT DOMAIN 888 915
CC FT DOMAIN 916 1239
CC FT DOMAIN 1315 1419
CC FT DOMAIN 1461 1766
CC FT NP_BIND 163 170
CC FT DOMAIN 644 666
CC FT MOD_RES 1734 1734
CC FT CONFLICT 1142 1142
CC SQ SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

Query Match
Best Local Similarity 21.8%; Pred. No. 46;
Matches 42; Conservative 33; Mismatches 60; Indels 58; Gaps 9;

QY 26 ESLSDFPDVMDYVPEPNEVFEESTQNLVKMLNCLSKQTKLGCCKVLPKLT 85
D 1561 EHCLTNFPLAEYROYLSDLAIQIY-----QQLVRYLENTL-----QPMIYSGMIE 1605
QY 86 QRIADVLRSLSTPEPCGLRGCVHVNLE---IENVCKRLD---RIVCDSSVVPFEL-- 136
D 1606 HETIOGV---SGVKPTGLRKRTSSIADEGTYLDSILRLQNSFHSVWCQCHGMDP--ELIK 1660
QY 137 -----TLVFKQENCSW-----TSFRDFEFGFGFSGFRRT--- 167
D 1661 QVVKOMFYIYGAITLNNLLRLKDMCSKQMOIRYNVSOLEEMLKDKNLNMSGAKETLEP 1720
QY 168 LILSSGFRLVKK 180
D 1721 LIGAAQLLVKKK 1733

RESULT 12
MYSA_MOUSE STANDARD; PRT; 1853 AA.
ID MYSA_MOUSE
AC 099104:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Va (Myosin 5a) (Dilute myosin heavy chain, non-muscle).
GN MYO5A OR DILUTE.
OS Mus musculus (mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=91141583; PubMed=1996138;
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RT "Novel myosin heavy chain encoded by murine dilute coat colour
RT locus.";
RL Nature 349:709-712(1991).
RN [2]
RP REVISIONS.
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RL Nature 352:547-547(1991).
CC -!- FUNCTION: MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR
CC ALTERNATIVELY, IT MAY BE REQUIRED FOR SOME POLARIZATION PROCESS
CC INVOLVED IN DENDRITE FORMATION.
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57377; CAA040651.1; -
DR PIR; A46761; A46761.
DR HSSP; P08799; ILVK.
DR MGD; MGI:105976; Myo5a.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEADV.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD003376; DIL; 1.
DR SMART; SM00015; IQ; 6.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 6.
KM Myosin repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Phosphorylation.
FT DOMAIN 1 765 MYOSIN HEAD-LIKE.
FT DOMAIN 766 788 IQ 1.
FT DOMAIN 789 813 IQ 2.
FT DOMAIN 814 836 IQ 3.
FT DOMAIN 837 861 IQ 4.
FT DOMAIN 862 884 IQ 5.
FT DOMAIN 885 913 IQ 6.
FT DOMAIN 914 1237 COILED COIL (POTENTIAL).
FT DOMAIN 1314 1443 COILED COIL (POTENTIAL).
FT DOMAIN 1685 1790 DILUTE.
FT NP_BIND 163 170 ATP (POTENTIAL).
FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1758 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 1853 AA; 215594 MW; 503E3D48CA6B766 CRC64;

```

```

DB 1585 ENCLTNEDELAERYVLSDLAIQIY-----QQLVRIENIL-----QPMIVSCMLE 1629
OY 86 QRIADQVLRSLSTPCGLRCVAMVNL-----IENCKRKLD-----RIYCDSSVPTFEL-- 136
DB 1630 HETIOGV-----SGVKPTGLRKRRTSSIADEGTYTLDSTLRQNSHRSVNCQGMGP--ELIK 1684
OY 137 -----TLVFKQENCSW-----TSFRDFEFSRGRFSSGFRRT--- 167
DB 1685 OYVKQFPIYGAITLNNLLLRKDCMSKGMQIRYVNSQLEEMLRDKNLNNSGAKETLEP 1744
OY 168 LIISGFRLYKK 180
DB 1745 LIQAQQLQYKK 1757

```

RESULT 13

```

ID MYSA_HUMAN STANDARD; PRT; 1855 AA.
AC Q9Y4I1; Q9UE30; Q9UE31; Q07902; Q16249; Q06053;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle)
DE (Myosin heavy chain 12) (Myoxin).
DE MYO5A OR MYH12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Skin.
RA Weuters B.H., Zimmermann R., Vosberg H.P.;
RT "The complete cDNA for human myosin heavy chain 12, a class V
RT myosin.";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CYS-1246.
RX MEDLINE=97351514; PubMed=9207796;
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RT "Griscelli disease maps to chromosome 15q21 and is associated with
RT mutations in the myosin-Va gene.";
RL Nat. Genet. 16:289-292(1997).
RN [3]
RP ERRATUM.
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RL Nat. Genet. 23:373-373(1999).
RN [4]
RP SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).
RC TISSUE=Fetal brain;
RX MEDLINE=94245227; PubMed=8188282;
RA Engle L.J., Kennett R.H.;
RT "Cloning, analysis, and chromosomal localization of myoxin (MYH12),
RT the human homologue to the mouse dilute gene.";
RL Genomics 19:407-416(1994).
RN [5]
RP SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=95136715; PubMed=7835087;
RA Moore K.J., Testa J.R., Francke U., Milatovich A., Copeland N.G.,
RA Jenkins N.A.;
RT "Cloning and regional assignment of the human myosin heavy chain 12
RT (MYH12) gene to chromosome band 15q21.";
RL Cytogenet. Cell Genet. 69:53-58(1995).
RN [6]
RP SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).
RA Edgar A.J., Bennett J.P.;
RT "Inhibition of dendrite formation in melanocytes transiently
RT transfected with antisense DNA to myosin V.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [7]

```

RP FUNCTION=99376094; PubMed=10448864;  
RA Menta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,  
RA Cheney R.E.;  
RT "Myosin-V is a processive actin-based motor.";  
RL Nature 400:590-593(1999).  
CC -I- FUNCTION: PROGRESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE  
CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.  
CC MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY  
CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE  
CC FORMATION.  
CC  
CC -I- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE  
CC CALMODULIN OR MYOSIN LIGHT CHAINS.  
CC  
CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC  
CC -I- DISEASE: DEFECTS IN MYO5A ARE A CAUSE OF GRISCELLI SYNDROME (GS).  
CC GS IS A RARE AUTOSOMAL RECESSIVE DISORDER THAT RESULTS IN  
CC CLUMPS OF PIGMENT IN HAIR SHAFTS, AND AN ACCUMULATION OF  
CC MELANOSOMES IN MELANOCYTES. MOST PATIENTS ALSO DEVELOP AN  
CC UNCONTROLLED T LYMPHOCYTE AND MACROPHAGE ACTIVATION SYNDROME,  
CC KNOWN AS HEMOPHAGOCYTIC SYNDROME, LEADING TO DEATH IN THE ABSENCE  
CC OF BONE MARROW TRANSPLANTATION.  
CC  
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC  
CC -I- SIMILARITY: CONTAINS 6 IQ DOMAINS.  
CC  
CC -I- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
DR EMBL: U090942; AACD00702.1; -;  
DR EMBL: Y077559; CAA69035.1; -;  
DR EMBL: Y077559; CAA69036.1; -;  
DR EMBL: Z229577; CAA80533.1; -;  
DR EMBL: S747997; AAB33211.1; -;  
DR EMBL: AF055459; AAC14188.1; -;  
DR HSSP: P08799; IMND.  
DR MIM: 160777; -;  
DR MIM: 214450; -;  
DR InterPro: IPRO02710; DIL.  
DR InterPro: IPR000048; IO.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF01843; DIL; 1.  
DR Pfam: PF00612; IO; 6.  
DR Pfam: PF00063; myosin\_head; 1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR ProDom: PD000355; myosinheavy; 1.  
DR SMART: SM003376; DIL; 1.  
DR SMART: SM00015; IO; 6.  
DR SMART: SM00242; MYSC; 1.  
DR PROSITE: PS50096; IO; 6.  
KW Myosin. Repeat. ATP-binding; Calmodulin-binding; Actin-binding;  
KW Coiled coil; Phosphorylation; Alternative splicing; Polymorphism.  
FT DOMAIN 1 765  
FT DOMAIN 766 788  
FT DOMAIN 789 818  
FT DOMAIN 814 836  
FT DOMAIN 837 861  
FT DOMAIN 862 883  
FT DOMAIN 885 914  
FT DOMAIN 914 1237  
FT DOMAIN 1237 1338  
FT DOMAIN 1338 1445  
FT DOMAIN 1445 1687  
FT NP\_BIND 163 170  
FT MOD\_RES 643 665  
FT MOD\_RES 1760 1760  
FT VASAPLIC 1321 1347  
FT VASAPLIC 1413 1413  
FT L--LFEEELVADPKKVOGEKRISLVYKRM1 (IN MISSING (IN ISOFORM 2)).

```

FT FT VARIANT 1246 1246 ISOFORM 3).
FT FT 668 668 R -> C.
FT FT CONFLICT 633 833 /FTID=VAR_010645.
FT FT CONFLICT 863 863 F -> L (IN REF. 1).
FT FT CONFLICT 922 922 MISSING (IN REF. 4).
FT FT CONFLICT 961 961 E -> G (IN REF. 1).
FT FT CONFLICT 1061 1061 H -> R (IN REF. 1).
FT FT CONFLICT 1089 1089 V -> L (IN REF. 5).
FT FT CONFLICT 1177 1177 E -> Q (IN REF. 4).
FT FT CONFLICT 1465 1477 D -> E (IN REF. 5).
FT FT CONFLICT 1471 1471 NIPREKDFQGM L -> SVLCACGCVTVR (IN REF. 4).
FT FT CONFLICT 1484 1484 K -> N (IN REF. 5).
FT FT CONFLICT 1855 AA: 215419 MW: 1C55AD57285FA9EC CRC64: E -> D (IN REF. 5).
SQ SEQUENCE 1855 AA: 215419 MW: 1C55AD57285FA9EC CRC64:

Query Match 7.7%; Score 77; DB 1; Length 1855;
Best Local Similarity 21.8%; Pred. No. 47;
Matches 42; Conservative 33; Mismatches 60; Indels 58; Gaps 9;

QY 26 ELSLSDFDYWDYVPEPINEVEFEESTICQNLVKMLKLECLSKSKQTKLGSKVLYPEKLT 85
DB 1587 EHCITNFEDLAEKROYLSLDAIQY-----QQLVRLVLENTL-----OPTVSGMLE 1631
QY 86 GRHADVLRLESTEGCLRGCVMAHNL-----IENVCKRLD--RIVCDSSVPTFEEL-- 136
DB 1633 HETIQCV---SGVKPTGLRKRTSTJADEQTYTLDSILQLNLFHVMKQHGMDP--ELIK 1686
QY 137 -----TLVEKQENCSW-----TSFNDFFPSRGFFSGFRRT-- 167
DB 1687 QYVKOMFYIIGAITNNLLLRKDCMSKQIRYRVNSQLEMLRDKNLINMSGAEFTLEP 1746
QY 168 ILTSGEFLYKKK 180
DB 1747 LIQAALLOVKKK 1759

RESULT 14
RPOB_MYCGA STANDARD: PRT: 1390 AA.
AC PA47715:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase
  beta chain) (RNA polymerase beta subunit).
GN RPOB.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RA BEADAEASHTVILL R.S.;
RA Skamrov A.V., Rozovskaya T.A., Goldman M.A., Feoktistova E.S.,
  Submitted (xxx-1995) to the EMBL/GenBank/DBJ databases.
RL -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
  OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
  SUBSTRATES.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
  [RNA](N).
CC -I- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
  ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
  BETA' CHAIN.
CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L38402; AAB40951.1; -
DR InterPro; IPR001572; RNA_pol_B.
DR Pfam; PF00562; RNA_pol_B.1.
DR PROSITE; PS01166; RNA_POL_BETA.1.
KW Transferrase; Transcription; DNA-directed RNA polymerase.
SQ SEQUENCE 1390 AA; 155997 MW; 1C02A893F073542 CRC64;

Query Match
Best Local Similarity 7.6%; Score 76; DB 1; Length 1390;
Matches 50; Conservative 36; Mismatches 91; Indels 20; Gaps 11;

QY 14 SILELDC-GYHPESLISFDYDYYV---PEPNINEVIFEESTQCNVKKME-NCLSK 67
DB 218 STITLLKAGLESEIKKEIFNNNDYLRSLSEFYEKOLINADIAQLIROESDRISK 277
QY 68 SKOTKLGCS-KYLVE--KLTORIADVLRLSTPECGLRGCVHVNLEIEN-VCKKLD 123
DB 278 VKSLPIDOKKMLVLDWMYKLNQ-EKOLLNSSNPPTKIESINTHTIGVLRKLCICKAK 336
QY 124 IYCDSSVVTFFELTVEFKQENCSWTS-FRDFEFSRGRF---SSG---FRRTILSSGF-- 174
DB 337 HVIQELISTRSUDNVAQKEEISTYOSILQHPFKRKRYLSSGKRHKVKKLRSERYQ 396
QY 175 RLVKKKLYSLIGTYYIE 191
DB 397 RTIADIKDLDCGNVIK 413

RESULT 15
GP37_HUMAN STANDARD; PRT; 613 AA.
ID GP37_HUMAN O15354; 014768; 000348;
AC 015354; 014768; 000348;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Probable G protein-coupled receptor GPR37 precursor (Endothelin B
DE receptor-like protein-1) (ETBR-LP-1).
GN GPR37.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97480717; PubMed=9339362;
RA Marezzi D., Golini E., Gallo A., Lombardi M.S., Matteoni R.,
RA Tocchini-Valentini G.P.;
RT "Cloning of GPR37, a gene located on chromosome 7 encoding a putative
RT G-protein-coupled peptide receptor, from a human frontal brain EST
RT library.";
RL Genomics 45:68-77(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Donohue P.J., Shapira H., Mantey S.A., Hampton L.L., Jensen R.T.,
RA Battey J.F.;
RT "A human gene encodes a putative G protein-coupled receptor highly
RT expressed in the central nervous system.";
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97289672; PubMed=9144577;
RA Zeng Z., Su K., Kyaw H., Li Y.;
RT "A novel endothelin receptor type-B-like gene enriched in the brain.";
RL Biochem. Biophys. Res. Commun. 233:559-567(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Jones K., Tin-Mollam A., Keppler D.;
RL Submitted (Jun-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ORPHAN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
```

```
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Y12476; CAA73080.1; JOINED.
DR EMBL: Y12477; CAA73080.1; JOINED.
DR EMBL: AF017262; AAB70008.1; -.
DR EMBL: U87460; AAC51281.1; -.
DR EMBL: AC004925; MAD08853.1; -.
DR MIM: 602583; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1.1.
DR PROSITE; PS00237; G_PROTEIN_REC_P1_1; FALSE_NEG.
DR PROSITE; PS00262; G_PROTEIN_REC_P1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT TRANSMEM 494 514
FT DOMAIN 515 531
FT TRANSMEM 532 552
FT DOMAIN 553 613
FT TRANSMEM 613 613
FT CHAIN 27 613
FT FT
FT DOMAIN 27 265
FT TRANSMEM 266 286
FT DOMAIN 287 299
FT TRANSMEM 300 320
FT DOMAIN 321 335
FT TRANSMEM 336 356
FT DOMAIN 357 379
FT TRANSMEM 380 400
FT DOMAIN 401 443
FT TRANSMEM 444 464
FT DOMAIN 465 493
FT
```

OY 152 DFFSRGRRSSGFRRLILSSGFLVKKKLYSLISTVI 190  
Db 468 KIRKAEKACTRGKRGKQIOLESQMCTVVALTLILYGCII 506

Search completed: October 11, 2002, 14:12:04  
Job time : 38 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: October 11, 2002, 14:02:42 ; Search time 49 seconds  
(without alignments)  
378.474 Million cell updates/sec

Title: US-09-881-636-2

Perfect score: 1002  
Sequence: 1 MVAWGSLSSKNPASISELD.....FRLVKKKLYSLIGTVIEGS 193

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90.5	9.0	1056	2	RIc1 protein - yea
2	85.5	8.5	711	2	hypothetical prote
3	85.5	8.5	1414	2	hypothetical prote
4	82	8.2	673	2	protein F25E5.7 [1
5	82	8.2	1002	2	conserved membrane
6	81.5	8.1	228	2	hypothetical prote
7	80	8.0	927	2	hypothetical prote
8	79.5	7.9	523	2	hypothetical prote
9	79.5	7.9	523	2	hypothetical prote
10	79.5	7.9	523	2	hypothetical prote
11	79	7.9	606	2	amiloride-sensitiv
12	79	7.9	549	2	cell death protein
13	79	7.9	1808	2	serine/threonine k
14	78.5	7.8	463	2	glycogen synthase
15	77.5	7.7	817	2	probable ATP-depen
16	77.5	7.7	1153	2	probable transposo
17	77.5	7.7	1584	2	hypothetical prote
18	77	7.7	394	2	protein F12M16.25
19	77	7.7	1417	2	TM916 ORF20 homolo
20	77	7.7	1828	2	Bloom's syndrome r
21	77	7.7	1830	1	myosin heavy chain
22	77	7.7	1833	1	myosin heavy chain
23	77	7.7	1855	1	myosin heavy chain
24	76.5	7.6	613	2	endothelin recepto
25	76.5	7.6	1309	2	protein T3P18.3 [1
26	76	7.6	1083	2	hypothetical prote
27	76	7.6	1871	2	probable DNA polym
28	76	7.6	1894	2	DNA-directed DNA p
29	75.5	7.5	963	2	hypothetical prote

30	75.5	7.5	1603	2	protein R10E8.6 [1
31	75.5	7.5	1696	2	hypothetical prote
32	75	7.5	509	2	hypothetical prote
33	75	7.5	1819	2	cag island protein
34	74.5	7.4	979	2	ubiquitin carboxyl
35	74.5	7.4	1171	2	hypothetical prote
36	74.5	7.4	1889	2	hypothetical prote
37	74.5	7.4	4196	2	dynein heavy chain
38	74	7.4	452	2	dynein heavy chain
39	74	7.4	585	2	cyclin cyc3 - alfa
40	74	7.4	1405	2	hypothetical prote
41	73.5	7.3	366	2	hypothetical prote
42	73.5	7.3	453	2	flagella-associate
43	73.5	7.3	453	2	hypothetical prote
44	73.5	7.3	949	2	hypothetical prote
45	73.5	7.3	1441	2	helicase II homo

## ALIGNMENTS

```

RESULT 1
S56039
RIc1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: Protein L1917; Protein YLR039C
C:Species: Saccharomyces cerevisiae
C>Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 20-Jun-2000
C:Accession: S56039; S64866
R:Mizuta, K.
submitted to the EMBL Data Library, December 1994
A:Reference number: S56039
A:Accession: S56039
A:Molecule type: DNA
A:Residues: 1-1056 <M12>
A:Cross-references: EMBL:D43895; NID:9619062; PID:9633123
R:Koetler, P.; Rose, M.; Entian, K.D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64863
A:Accession: S64866
A:Molecule type: DNA
A:Residues: 1-1056 <KOE>
A:Cross-references: EMBL:D73211; NID:91360365; PID:91360366; MIPS:YLR039C
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:RIC1
A:Cross-references: SGD:S0004029; MIPS:YLR039C
A:Map position: 12R
C:Superfamily: Saccharomyces cerevisiae RIC1 protein

Query Match          9.08; Score 90.5; DB 2; Length 1056;
Best Local Similarity 25.08; Pred. No. 4;
Matches 42; Conservative 23; Mismatches 52; Indels 51; Gaps 8;

OY 24 HPESLSDFDWDVYVPEPNLEVEFEESTQNLVKNLENCISKSKOTKICGSKVLPEK 83
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 613 NPVELSDYVMEVNPPE---NTIYF-----SLAVNTLSRKRLKLNKSKNNNAEK 660
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 84 L-----TORIADVLRLSSTPECGIRG-----VMHVNLEIENCKKIDR 123
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 661 QPDALLKTAETILLVDTQTIVFDV--ISTVHPGCLNIIKFEQYLKINIFDIVLPNKIEM 718
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 124 IVCSSVVPFELTLVFEQKNCSSWTSFRD--FFFSGRSSGFRRLTL 170
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 719 II-----NMKEGLLF-----FADRRFKIKGVKGWQDTLTL 750
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 2
T30107
hypothetical protein F4A4.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30107
R:Geisels, C.; Bradshaw, H.

```



OY 138 -----LVFKQENCWTSFRDFEFGSSGFR-----RTLLSS---GFRLL 176  
 Db 797 FLGVISEPIVFDNK-----MPDLFENKERYADVFKFEYKKMNTVLTNTQNVAGFSW 850  
 OY 177 VKKKLYSLGTTV 189  
 Db 851 INDKRIAFIASGI 863

## RESULT 6

B90292  
 hypothetical protein SSO1357 [imported] - Sulfolobus solifarius  
 C:Species: Sulfolobus solifarius  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: B90292  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayer, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.  
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solifarius complete genome.  
 A:Reference number: A9139  
 A:Accession: B90292  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-228 <KUR>  
 A:Cross-references: GB:AE006641; MID:g13814569; PIDN:AAK41593.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SSO1357

Query Match 8.1%; Score 81.5; DB 2; Length 228;  
 Best Local Similarity 22.8%; Pred. No. 4.2; Mismatches 48; Indels 33; Gaps 5;  
 Matches 31; Conservative 24;

OY 53 TCQNLVKMLNCLSKSKQTKLGSKVLYPE-KLTQRIADVLRLSTPEPCGLGCMVHN 111  
 Db 17 TCKESLKM-----SVIIPAFNEERRIGTKLTKISSTLP-----N 51  
 OY 112 LEIENVCKKLDRIVCDSVVPFELTVFKQENCWTSFRDFEFGSSGFRRTLLS 171  
 Db 52 AEVVAVFDGHDN--TPEVVKRFPVKLIISKRLG---KGMALKGITSNFORVLLD 104  
 OY 172 SGFRLVKKLYSLIGT 187  
 Db 105 ADFPTEELNKILST 120

## RESULT 7

T00357  
 hypothetical protein KIAA0685 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1998 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00357  
 R:ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kocani, H.; Nomura, N.  
 DNA Res. 5, 169-176, 1998  
 A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete  
 A:Reference number: Z14142; MUID:98403880  
 A:Accession: T00357  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-927 <ISH>  
 A:Cross-references: EMBL:AB014585; MID:g3327183; PIDN:EAA31660.1; PID:g3327184  
 A:Experimental source: brain  
 C:Genetics:  
 A:Note: KIAA0685

Query Match 8.0%; Score 80; DB 2; Length 927;  
 Best Local Similarity 32.0%; Pred. No. 32; Mismatches 53; Indels 16; Gaps 7;  
 Matches 41; Conservative 18;

OY 10 KNPAISIELDLCGY-----HPESLSDFYWDYVVEPNLNEYI--FEESTCONLY- 58  
 Db 76 KYPNTACELTCDVPQISDRILGGDESTLTL--YDFLDHEPPLNPLASFSFKTIGNLTA 133

OY 59 KMLNCLSKSKQTKLGSKVLYPEKLTQRIADVLRL-STPEPCGLGCMVHNEIENY 117  
 Db 134 RKTEQYITFLKKDKRKTISLVL-KHIGTSALMDLLALVSCVEPAGLRQVLDH-LNEENY 191  
 OY 118 CKKLDRIY 125  
 Db 192 IORLVELI 199

## RESULT 8

D85538  
 hypothetical protein 20521 [imported] - Escherichia coli (strain O157:H7, substrain E  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: D85538  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May-  
 Miller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda-  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: AB5480; MUID:21074935; PMID:11206551  
 A:Accession: D85538  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-523 <STO>  
 A:Cross-references: GB:AE005174; MID:g12513274; PIDN:AAG54768.1; GSPDB:GN00145; UNCP:  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: 20521

Query Match 7.9%; Score 79.5; DB 2; Length 523;  
 Best Local Similarity 24.5%; Pred. No. 18; Mismatches 62; Indels 37; Gaps 7;  
 Matches 40; Conservative 24;

OY 29 LSDFDYWDYVVEPNLNEYIFEESTCONLYKMLNCLSKSKQTKLGC--SKVL--VPER 83  
 Db 63 LSVLANNNSALTPPNFN--IDATSDLDYIKIDFDLSPKQKTYLCCFWNKIASSLPER 120  
 OY 84 LTORIADVLRLSTPEPCGLGCMVHN-----LEIF-----NVCK 119  
 Db 121 YNSTIKHNIIFYKDGENLMIKPTISIVNEVVKYTSPIEDKDNGYDFSGLYLAHSNIG 180  
 OY 120 KLDRIVCDSVVPFELTVFKQENCWTSFRDFEFGSSGFRSS 162  
 Db 181 K-----DPKDPDIDFGIDMNCNCNVNHEHTFYGVAFYTN 217

## RESULT 9

H90687  
 hypothetical protein ECS0472 [imported] - Escherichia coli (strain O157:H7, substrain  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: H90687  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
 A:Reference number: A9629; MUID:21156231; PMID:11238796  
 A:Accession: H90687  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-523 <HAV>  
 A:Cross-references: GB:BA000007; PIDN:BA03895.1; PID:g13359923; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECS0472

Query Match 7.9%; Score 79.5; DB 2; Length 523;  
 Best Local Similarity 24.5%; Pred. No. 18; Mismatches 62; Indels 37; Gaps 7;  
 Matches 40; Conservative 24;

OY 29 LSDFDYWDYVVEPNLNEYIFEESTCONLYKMLNCLSKSKQTKLGC--SKVL--VPER 83  
 Db 63 LSVLANNNSALTPPNFN--IDATSDLDYIKIDFDLSPKQKTYLCCFWNKIASSLPER 120

Db 63 LSVLNTNSALPPDPNFN--IDATSDLSYILKDPDRLSPKQKQTLTCCFWMKIASLPEP 120

OY 84 LTORAADOVLNLSPEPGGLRGCMVHN-----LELE-----NVC 119

Db 121 YNSTKKNHIIYKCGENLMIGTISIVNEVKYTSLEPKDNGYIDPSGLYLAHSING 180

OY 120 KLDRIYCDSSVPTPELTLVFKQENCSWTSPRDFEFSNGRFS 162

Db 181 K-----DPNKDPDIDFGIDMGNCNSVNVNEHHYFYGVKPEFN 217

Db 715 SLIPONOKDTHIKIGRLLQNT 737

## RESULT 13

C70363

glycogen synthase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 18-Jun-1999

C:Accession: C70363

R:Decker, G.; Warren, P. V.; Gaasterland, T.; Young, W. G.; Lenox, A. L.; Graham, D. E.; O'V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:9819666

A:Accession: C70363

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-463 <AOF>

A:Cross-references: GB:AE000704; NID:g2983301; PIDN:AAC06894.1; PID:g2983303; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: glgA

C:Superfamily: starch synthase

Query Match 7.8%; Score 78.5; DB 2; Length 463;  
Best Local Similarity 19.6%; Pred. No. 19;  
Matches 37; Conservative 37; Mismatches 78; Indels 37; Gaps 5;

QY 1 MWATGSLSSKNPASISETL-LDCGYHPESILSDPDY-----WDYVPEPNEVIRE 50

DB 198 IVFSDLITVSPFYAKKEIOTQTEYGLGECVLAKKSYKRLGLNGIDYEVWNEKRYIQ 257

QY 51 ESTCONLYKMLENCLSKSKQKRLGCSKVLP--EKLTORIADVLRSLSTPECGLRGCV 108

DB 258 NVSLNYSKKEFKNEFLSELGIEAEKPLISFINFTTHQKGYEL----- 301

QY 109 HYNLEIENWCKLDRIYCDSSVVPFELTLVFKQENCSWTSRDPFSGRGSSGPRRL 168

DB 302 -----ILMCAEEMKLMNMFVFLGTGEYENAFLDVSKYIKNKVF--AEFNEGPARL 352

QY 169 ILSSGFRLV 177

DB 353 YASSDFILM 361

## RESULT 14

T04325

probable ATP-dependent proteinase LON2 (EC 3.4.21.-), mitochondrial - maize

C:Species: Zea mays (maize)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 19-Jan-2001

C:Accession: T04325

R:Barakat, S.; Pearce, D. A.; Sherman, F.; Rapp, W. D.

Plant Mol. Biol. 37, 141-154, 1998

A:Title: Maize contains a Lon protease gene that can partially complement a yeast *pim1*-C

A:Reference number: Z15282; MUID:98281582

A:Accession: T04325

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-964 <BAR>

A:Cross-references: EMBL:085495; NID:g1816587; PIDN:AAC50021.1; PID:g1816588

A:Experimental source: strain B73

C:Genetics:

A:Gene: LON2

C:Function:

A:Description: serine proteinase

C:Superfamily: ATP-dependent serine proteinase La

C:Keywords: ATP; DNA binding; hydrolase; mitochondrial matrix; mitochondrion; molecular

F:455-462/Region: nucleotide-binding motif A (P-loop)

F:518-523/Region: nucleotide-binding motif B

F:461/Binding site: ATP (lys) #status predicted

F:863/Active site: Ser #status predicted

Query Match 7.8%; Score 78.5; DB 1; Length 964;

Best Local Similarity 27.6%; Pred. No. 46;  
Matches 34; Conservative 28; Mismatches 42; Indels 19; Gaps 8;

QY 5 GSLSKNPAS-ISELLDCGYHPESILSDPDYVVPFELTLVFKQENCSWTSRDPFSGRGSSGPRRL 62

DB 528 GKHSDDPASPALLELD-----PEQNVNPLDH--YLDVPIJLSKVLK---VGTANVTEKIP 578

QY 63 NCLSKSKQKRLGCSKVLP-EKLTORIAQVLRSLSTPECGLRGCVHVN-----LEIENV 117

DB 579 NPL-LDRMELIAGVITDEKM--HIAVDYLEKNTQACGIRPEQVEYVDTLLALIEHY 635

QY 118 CKK 120

DB 636 CRE 638

## RESULT 15

D85049

probable transposon protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C:Accession: D85049

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: D85049

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-817 <STO>

A:Cross-references: GB:NC\_001268; NID:g7267145; PIDN:CAH80813.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4G03900

A:Map position: 4

Query Match 7.7%; Score 77.5; DB 2; Length 817;  
Best Local Similarity 20.1%; Pred. No. 47;  
Matches 36; Conservative 29; Mismatches 71; Indels 43; Gaps 5;

QY 7 LSSKNPASISELLDCG-----YHPESILSDPDYVVPFELTLVFKQENCSWTSRDPFSGRGSSGPRRL 53

DB 540 LASVNPRTKTDYGVNGCHEKKMRGCGEKNHMHKESIMELSYWKDILRNINIDVHNEKNF 599

QY 54 CONLYKMLENCLSKSKQKRLG-----CSK-----VLVEKLTQRIADVLR 95

DB 600 LDINIMFTLLGVKSKSDNIMSRDIEKYSRGLHIDYGRAPFPYKLTLEAKOSLQF- 658

QY 96 SSTEPGLRGCVHVNLEIENWCKLDRIYCDSSVVPFELTLVFKQENCSWTSRDPFSGRGSSGPRRL 154

DB 659 -----CYKH-DVREPDGKSHDCVFMERLLPFIFAELDRNVHLALSGINAF 706

Search completed: October 11, 2002, 14:14:48

Job time : 52 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 12:29:47 ; Search time 63 Seconds  
(without alignments)  
340.274 Million cell updates/sec

Title: US-09-881-636-2

Perfect score: 1002

Sequence: 1 MVAATGSLSSKNPASPISLDD.....FRLVKRLYLIGTTVIEGS 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Database : Listing first 45 summaries

1: A\_Geneseq\_032802.\*  
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	272.5	27.2	177	20	AAV59820 Human normal ovar
2	271	27.0	232	22	AAV03635 Hypoxia-regulated
3	271	27.0	232	22	AAE03922 Human gene 25 enco
4	271	27.0	233	21	AA851661 Human secreted pro
5	261	26.0	229	20	AAV03634 Hypoxia-regulated
6	211.5	21.1	299	22	AB863344 Drosophila melanog
7	193	19.3	280	22	AB863379 Drosophila melanog
8	112.5	11.2	74	21	AA845148 Human secreted pro
9	85	8.5	1066	22	AB864251 Drosophila melanog
10	82	8.0	934	22	AA860490 Human cell cycle a
11	80.5	8.0	498	22	AA892900 Human protein sequ

12	79.5	7.9	606	22	AB859933 Drosophila melanog
13	79.5	7.9	860	22	AAW79768 Human protein SEO
14	79	7.9	169	22	AAU20460 Human secreted pro
15	79	7.9	261	15	AA853277 Ced-4 (D262X). Ca
16	79	7.9	315	21	AA843007 Human ORF27771
17	79	7.9	315	21	AA843074 Human ORF27771
18	79	7.9	322	22	AB811760 A human a neuronal
19	79	7.9	409	15	AA853278 Human voltage gate
20	79	7.9	549	15	AA853276 Ced-4 (W401X(sic))
21	79	7.9	549	15	AA847465 Ced-4. Caenorhabd
22	78	7.8	179	22	AAU31632 Novel human secret
23	78	7.8	224	22	AB805646 Novel human diagno
24	78	7.8	224	22	AB805652 Novel human diagno
25	78	7.8	371	22	AAU28240 Novel human secret
26	78	7.8	1056	22	AB815548 Novel human diagno
27	78	7.8	1056	22	ABG09490 Novel human diagno
28	77.5	7.7	613	21	AAV90667 Human mutant G pro
29	77	7.7	515	18	AAW31547 Bloom's syndrome B
30	77	7.7	739	18	AAW31549 Bloom's syndrome B
31	77	7.7	1417	18	AAW31548 Bloom's syndrome B
32	77	7.7	1417	18	AAW31550 Bloom's syndrome B
33	77	7.7	1417	18	AAW31551 Bloom's syndrome B
34	77	7.7	1418	18	AAW15264 Bloom syndrome act
35	76.5	7.6	314	21	AA86375 Protein encoded by
36	76.5	7.6	613	17	AA898261 Human endothelin-b
37	76.5	7.6	613	19	AAW37799 Amino acid sequenc
38	76.5	7.6	1262	22	AB859586 S.cerevisiae apopt
39	76	7.6	1364	22	AA870912 Human G protein-co
40	75.5	7.5	348	22	ABG05997 Amino acid sequenc
41	75.5	7.5	613	21	AAV90637 Polypeptide sequen
42	75	7.5	446	20	AAV36840 A human voltage ga
43	75	7.5	748	22	AAU07878 Human voltage gate
44	74.5	7.4	323	21	AAV84376
45	74.5	7.4	325	22	AB811805

## ALIGNMENTS

RESULT 1	
AAV59820	
ID	AAV59820 standard; Protein; 177 AA.
AC	AAV59820;
DT	18-JAN-2000 (first entry)
DE	Human normal ovarian tissue derived protein 97.
XX	Human; ovary; screening; ovarian cancer; treatment.
KW	
XX	
OS	Homo sapiens.
XX	
PN	DE19816395-A1.
XX	
PD	07-OCT-1999.
XX	
PF	03-APR-1998; 98DE-1016395.
XX	
PR	03-APR-1998; 98DE-1016395.
XX	
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.
XX	
PI	Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX	
DR	WPI: 1999-552352/47.
DR	N-PSDB; AA241261.
PT	Nucleic acid sequences potentially useful in diagnosis or therapy of
XX	ovarian cancer -
PS	Claim 23; Page 255; 274ppp; German.
XX	





DE Hypoxia-regulated gene RTP801 product.  
 XX  
 KW Hypoxia-regulated gene; therapeutic; diagnostic; hypoxia; ischemia;  
 KM apoptosis; angiogenesis; tumorigenic cell; trauma; limb reattachment;  
 KW revascularisation.  
 XX  
 OS Rattus sp.  
 XX  
 PN W09909049-A1.  
 XX  
 PD 25-FEB-1999.  
 XX  
 PF 21-AUG-1998; 98MO-US17296.  
 XX  
 PR 21-AUG-1997; 97US-0056453.  
 XX  
 PA (KOHN/) KOHN K I.  
 XX (QUAR-) QUARK BIOTECH INC.  
 PI Elnat P, Skalter R;  
 XX  
 DR WPI; 1999-180965/15.  
 DR N-PSDB; AAX29137.  
 XX  
 PT New isolated hypoxia-related genes - used to develop products for  
 PT use in therapy and diagnosis in e.g. hypoxia, ischemia, apoptosis  
 PT and angiogenesis  
 XX  
 PS Claim 14; Page 81-82; 92pp; English.  
 XX  
 CC Sequences AAY03632-36 represent products of hypoxia-regulated genes of  
 CC the invention. The genes and their products can be used therapeutically  
 CC and diagnostically in hypoxia, ischemia, apoptosis and angiogenesis. The  
 CC products and methods can be used for e.g. inducing apoptosis in  
 CC tumorigenic cells or angiogenesis in trauma situations where e.g. a limb  
 CC must be reattached or in a transplant where revascularisation is needed.  
 XX  
 SQ Sequence 229 AA:  
 Query Match 26.0%; Score 261; DB 20; Length 229;  
 Best Local Similarity 30.7%; Pred. No. 2.8e-21;  
 Matches 65; Conservative 45; Mismatches 68; Indels 34; Gaps 5;  
 QY 3 ANGSSSKKPA-----SISELDC-----GIHPESLISDPD 33  
 DB 12 SSSSSSRTPADRPBSAAGSAREEGIDRCASLSDCESLSDSSNSGFGPE--EDSS 68  
 QY 34 YMDYV-VPEPNTNEVFEESTCONLYKMLENCISKSKOTKLGSKYLVPEKLTQRIADQY 92  
 DB 69 YLDGVSILPPELLSDPEDHLCANLMQLOESLSQARLGRPARLMLPSQLLSQYKEL 128  
 QY 93 LRLISTPCGLRGCVAHVNLIEINWCKKLDRIYCDSSVVPFELTLVFKQENCSWTSFRD 152  
 DB 129 LRLAYSEPCGLRGALDVCVEQKSCSHVAQALADPSLVPFTQTLVLRLDSLMLPKIGD 188  
 QY 153 FFFS-RGRFSSGFRRTLISSGFRVLVKKLYS 183  
 DB 189 LLSANSSLVPGYSOSLTSTGFRVTKKLYS 220  
 RESULT 6  
 ABB63344  
 ID ABB63344 standard; Protein; 299 AA.  
 XX  
 AC ABB63344;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 16824.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX

OS Drosophila melanogaster.  
 XX  
 PN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL07447.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure: SEQ ID NO 16824; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 299 AA:  
 Query Match 21.1%; Score 211.5; DB 22; Length 299;  
 Best Local Similarity 35.9%; Pred. No. 1.7e-15;  
 Matches 51; Conservative 24; Mismatches 60; Indels 7; Gaps 3;  
 QY 43 NINEVFEESTCONLYKMLENCISKSKOTKLGSKYLVPEKLTQRIADQYLRISSTFPG 102  
 DB 162 NLDDV--SASVAVELISQQLQALDRDAKRRLHACTEVTLPNDLQRIAAELIRMSEREPCG 219  
 QY 103 LRGCVAHVNLIEI-NVCKKLDRIYCDSSVVPFELTLVFKQENCSWTSFRDFFSRRGFS 161  
 DB 220 ERACITLFIIEFESPPNKVRYKRVDPDYSIFELVTLRLQDKSGMSLVPPQPIKRLTRS 279  
 QY 162 SGFRRTLISSGFRVLVKKLYS 183  
 DB 280 N---TINISPDFTLTKKLYS 297  
 RESULT 7  
 ABB63379  
 ID ABB63379 standard; Protein; 280 AA.  
 XX  
 AC ABB63379;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 16929.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.

XX		23-MAR-2001; 2001WO-US092231.
PE		
XX		23-MAR-2000; 2000US-191637P.
PR		
XX		11-JUL-2000; 2000US-0614150.
PA	(PEKE ) PE CORP NY.	
XX		
PI	Venter JC, Adams M, Li PWD, Myers EW;	
DR	WPI: 2001-656860/75.	
N-PSDB:	ABL07482.	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -	
PS	Disclosure: SEQ ID NO 16929; 21pp + Sequence Listing; English.	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABB57737-ABB72072).	
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
CC		
XX	Sequence 280 AA;	
SQ		
Query Match	19.3%; Score 193; DB 22; Length 280;	
Best Local Similarity	29.2%; Pred. No. 2e-13;	
Matches	49; Conservative 36; Mismatches 61; Indels 22; Gaps 5;	
OY	30 SDFPDVMDVVPDE-----PNINEVIFEESTCNLVKMLENCLSKSQTGLCSKYVL 79   : : :   : : :   : : :   : : :   : : :   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	99 SNSNYYYAAADEEGSGADYALSNYDKRAVELLS---LRILDE-LRAKSRLHCTEYS 153   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
OY	80 VPEKLTQLAADVLELSTSPGCLGCVMHVALEIE-NCKKLDRIYCDSSVYPFELTL 138   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	154 LPCDLTPSVARKETIVSEKEPRGICGTIYIEFEDEPKNSRINASIKVDSDTVSFEEYVL 213 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
OY	139 VFKOCSMTSEFRDFFFSGRFPSSGFRRLLISGFRLLVKKKLYSLIG 186 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	214 TLROQHRCGWTSLLP-----QFMKSLARTITISPYYTITKNKLVSADG 255 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
RESULT 8		
AAB45148		
ID	AAB45148 standard; Protein: 74 AA.	
XX	AAB45148:	
XX		
DT	12-FEB-2001 (first entry)	
DC		
XX	Human secreted protein sequence encoded by gene 29 SEQ ID NO: 89.	
KM	Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic; cerebroprotective; neurotropic; neuroprotective; antibacterial; vitriptide; fungicide; cancer; opthalmoalological; autoimmune disease; hyperproliferative disorder; cardiovascular disorder; cerebrovascular disorder; wound healing; nervous system disorder; aging; chemotaxis.	
OS	Homo sapiens.	
XX		
XX	WO2000058467-A1.	
XX		
DD	05-OCT-2000.	

[illegible]

PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR MPI: 2001-656860/75.  
DR N-PSDB; ABL08354.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure: SEQ ID NO 19545; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB16351), expressed DNA  
CC sequences (AB161840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1066 AA:  
  
Query Match 8.5%; Score 85; DB 22; Length 1066;  
Best Local Similarity 25.6%; Pred. No. 2.6;  
Matches 30; Conservative 21; Mismatches 44; Indels 22; Gaps 4;  
  
QY 36 DYVPPNNINVEYFEESTCONLYKMLENLSKSKOTKLGCKYLVPEKLTORIAQDVRL 95  
DB 648 NYLAVYVPIQHFV-----VQALVSL-----AKLTKYGFDSYKKEVNFQNLDEYKKR 696  
  
QY 96 STEPCGLRGCVNHNLEIENCKKIDRIYCDSSVPTFELTVFKQENCSWTSFSD 152  
DB 697 -----LOGSYEHCTIGV-----ILSLVCEKNSVEMDVQVFSKMKRIATSFSD 742  
  
RESULT 10  
AAB60490  
ID AAB60490 standard; Protein; 934 AA.  
XX  
AC AAB60490;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Human cell cycle and proliferation protein CCYPR-38, SEQ ID NO:38.  
XX  
KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
KW antagonist; gene therapy; detection; gene therapy;  
KW transgenic animal disease model; immune disorder;  
KW developmental disorder; cell signalling disorder;  
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
KW menstrual cycle disorder; bacterial infection.  
XX  
OS Homo sapiens.  
XX  
PN WO200107471-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 21-JUL-2000; 2000WO-US19948.  
XX  
PR 21-JUL-1999; 99US-0145075.  
PR 08-SEP-1999; 99US-0153129.

PR 10-NOV-1999; 99US-0164647.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
XX  
DR MPI: 2001-112727/12.  
DR N-PSDB; AAF59627.  
XX  
PT Human cell cycle and proliferation proteins and polynucleotides are  
PT used to treat, diagnose and prevent immune, developmental and cell  
PT signalling disorders and cell proliferative disorders including cancer -  
XX  
PS Claim 1; Page 146-148; 205pp; English.  
XX  
CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and  
CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.  
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
CC associated with decreased expression of functional CCYPR, while CCYPR  
CC antagonists are used to treat diseases or conditions associated with  
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
CC that specifically bind to CCYPR, and in drug screening methods to  
CC identify compounds that modulate the activity of CCYPR. CCYPR  
CC nucleotides can be used to generate transgenic animal models of human  
CC disease, and can be used in gene therapy in target cells with genetic  
CC abnormalities with respect to the expression of CCYPR for the  
CC treatment or prevention of a disorder associated with CCYPR.  
CC Diseases which can be diagnosed, treated and prevented using CCYPR  
CC proteins, nucleic acids, agonists or antagonists include immune,  
CC developmental and cell signalling disorders, and cell proliferative  
CC disorders including cancer. Specific examples of these disorders  
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
CC diabetes mellitus, disorders of the menstrual cycle and infections  
CC caused by bacteria.  
XX  
SQ Sequence 934 AA:  
  
Query Match 8.2%; Score 82; DB 22; Length 934;  
Best Local Similarity 25.4%; Pred. No. 4.7;  
Matches 53; Conservative 29; Mismatches 75; Indels 52; Gaps 10;  
  
QY 10 KNPASISELDGCGY-----HPESLSDFDYDWDYVPPNNINVEY--FEESTCONLY- 58  
DB 76 KYPTNACELLTCVPOISDRLGDESLSL--YDFLDHPPNLPNLLASFSTKTLNLA 133  
  
QY 59 KMLENLSKSKOTKLGCKYLVPEKLTORIAQDVRL--STEPGCLRGCVNHNLEIENY 117  
DB 134 RKTBOVITFLKKKDFISVL-KHIGTSALMDLLRLVSCVEPAGKQDVLAH-LNEEKV 191  
  
QY 118 CKKIDRIY-----CD-----SSVPTFELTVFKQENCSWT 148  
DB 192 IQRVLVEIHRSQEDENSNASOTLCIDIVKRGROGSGLOALPDPDLTLAESRQDCEV 251  
  
QY 149 SFRDFFFSRGRFSSGFRRTLLISSGFRVL 177  
DB 252 LTKNMF-----DGRTECSLVSGTVL 273  
  
RESULT 11  
AAB92900  
ID AAB92900 standard; Protein; 498 AA.  
XX  
AC AAB92900;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:11520.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.  
XX PN EPI074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602  
XX PT full-length cDNAs defined in the specification, and for the detection  
XX PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX PT full-length cDNAs -  
XX  
XX PS Claim 8; SEQ ID 11520; 2537bp + CD ROM; English.  
XX  
XX CC The present invention describes primer sets for synthesizing 5602  
XX CC full-length cDNAs defined in the specification, where a primer set  
XX CC comprises (a) an oligo-dT primer and an oligonucleotide complementary  
XX CC to the complementary strand of a polynucleotide which comprises one of  
XX CC the 5602 nucleotide sequences defined in the specification, where the  
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX CC of an oligonucleotide comprising a sequence complementary to the  
XX CC complementary strand of a polynucleotide which comprises a 5'-end  
XX CC sequence and an oligonucleotide comprising a sequence complementary to a  
XX CC polynucleotide which comprises a 3'-end sequence, where the  
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of  
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in  
XX CC the specification. The primer sets can be used in antisense therapy and  
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX CC particularly full-length cDNAs. The primers are also useful for the  
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by  
XX CC the full-length cDNAs. The primers allow obtaining of the full-length  
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX CC represent oligonucleotides, all of which are used in the exemplification  
XX CC of the present invention.  
XX  
XX SQ Sequence 498 AA;  
SQ  
Query Match 8.0%; Score 80.5; DB 22; Length 498;  
Best Local Similarity 23.1%; Pred. No. 2.9;  
Matches 39; Conservative 33; Mismatches 58; Indels 39; Gaps 7;  
OY 27 SLSDPDYWDYVDE--PDLNEVFEESTCONLYKMLENCSKSKOTKLGSKVLVPERL 84  
DB 292 NLMNIRDMNKTWVLELCGINELSHPRNLMLVQLVPMNMTSRGCLR-QCLSLVIYIKL 350  
OY 85 TORIAODVLRISTPERGRLGCVNH--VNLIEIENVCKL-----DRIVCS----- 128  
DB 351 LDEKHEDVPANSN-----LQVSVLHRYLVOMKPSDLKKVKKKAEDPDGIDDSLHLE 405  
OY 129 ----SVVPTFELTLVFKQENCSWTSFRDFEFSRGSFGFRRLIISG 173  
DB 406 LEKQAYVLTLYLLHLVGEVSCSHS-----FSSGQRKHRYLLAG 443

RESULT 12  
AAB59933

ID AAB59933 standard; Protein; 606 AA.  
XX AC AAB59933;  
XX 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 6591.  
XX KW Drosophila: developmental biology; cell signalling; insecticide;  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PMD, Myers EW;  
XX DR N-PSDB; ABL04036.  
XX WPI; 2001-656860/75.  
XX  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX PT interactions -  
XX  
XX PS Disclosure; SEQ ID NO 6591; 21bp + Sequence Listing; English.  
XX  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signalling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX CC sequences (AAB57737-AAB72072), and the encoded proteins  
XX CC (AAB57737-AAB72072).  
XX CC The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 606 AA;  
SQ  
Query Match 7.9%; Score 79.5; DB 22; Length 606;  
Best Local Similarity 19.3%; Pred. No. 4.9;  
Matches 35; Conservative 32; Mismatches 47; Indels 67; Gaps 9;  
OY 37 YVVPENLNEVI-----FEESNCONLYKMLENCSKSKOTKLGSK 77  
DB 382 YVSTNDNLHEITPERKQCLFDDERSIRFRRSISQSNCR-----ECLANYVSKGCGAK 435  
OY 78 VLVPEKLRQRIADVLRISTPERGCLR--GCV-----MHVNLIEIENVCKLDR--IYC 126  
DB 436 FWMKPRPL-----GTPVCGIKINDICYTSAODELYTLQMQNTYAKSIDESVDIIC 483  
OY 127 D-----SSVPTFELTLVFKQENCSWTSFRD-----FEFSRGSFGFRRLI 168  
DB 484 NCMRACTSLEYNEFISRAKYDVAKTIRAREYEHTDAIGSLSYFKEHPTA-IKRTI 542  
OY 169 I 169  
DB 543 L 543

RESULT 13  
AAM79768  
ID AAM79768 standard; Protein; 860 AA.

[illegible]

Db	753	VCRFENLNCNCH---	IDNQGVEDIYRDMTKVAHVQY-?	-----	WHEFPGAMF	800
Oy	151	RDFEFSNGRPFSSGFRRTLISGFRVLVKKL-----	YSLIGTVIEGS	193		
Db	801	RKMPALGGEKTLIRVIIVIGVITLLPRLLPYLLQMIKSFYATLVYONA	850			
RESULT 14						
AAU20460						
ID	AAU20460	standard; Protein; 169 AA.				
XX	AAU20460:					
AC						
XX						
XX	04-DEC-2001	(first entry)				
DE						
XX		Human secreted protein, Seq ID No 452.				
XX						
RW		Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;				
RW		rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;				
RW		cerebroprotective; thrombolytic; antimicrobial; ophthalmological;				
RW		cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;				
RW		multiple sclerosis; cancer; hyperproliferative disorder; infection;				
RW		Gaucher's disease; neurological disease; cerebrovascular disorder;				
RW		thrombosis; wound healing.				
OS		Homo sapiens.				
PN	WO200155326-A2.					
PD	02-AUG-2001.					
XX						
PF	17-JAN-2001; 2001WO-US01347.					
XX						
PR	31-JAN-2000; 2000US-0179065.					
XX						
PA	(HUMA-) HUMAN GENOME SCI INC.					
XX						
PL	Rosen CA, Barash SC, Ruben SM;					
DR	WPI: 2001-451931/48.					
DR	N-PSDB; AAS33169.					
XX						
PT	New nucleic acids and polypeptides, useful for diagnosing, preventing					
PT	or treating medical conditions -					
XX						
PS	Claim 11; SEQ ID NO 452; 753bp; English.					
XX						
CC	The invention relates to novel isolated nucleic acid molecules (I)					
CC	encoding human secreted proteins (II). (I) and (II) are used to prevent,					
CC	treat or ameliorate a medical condition in e.g. humans, mice, rabbits, in					
CC	goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in					
CC	the prevention, treatment and diagnosis of diseases associated with					
CC	inappropriate expression of secreted proteins. (I) and complementary					
CC	sequences may also be used as DNA probes in diagnostic assays (e.g.					
CC	polymerase chain reactions (PCR)) to detect and quantitate the presence					
CC	of similar nucleic acid sequences in samples, and so which patients may					
CC	be in need of restorative therapy. (II) may also be used as antigens in					
CC	the production of antibodies and in assays to identify modulators					
CC	(agonists and antagonists) of the expression and activity of the secreted					
CC	proteins. The anti-(II) antibodies and antagonists may also be used to					
CC	down regulate expression and activity of (II). The anti-(II) antibodies					
CC	may also be used as diagnostic agents for detecting the presence of (II)					
CC	in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The					
CC	disorders include for example: immune/autoimmune diseases (e.g. HIV					
CC	(human immunodeficiency virus) infections, anaemia, rheumatoid arthritis					
CC	and multiple sclerosis), cancers and hyperproliferative disorders (e.g.					
CC	melanoma, neoplasms of the breast or liver, Sezary syndrome and					
CC	Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,					
CC	Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/					
CC	cerebrovascular disorders (e.g. cardiac arrest, tachycardia,					
CC	angina and thrombosis), infections caused by bacteria, viruses and					
CC	fungi and ocular disorders (e.g. corneal infections). (I) and (II),					

CC agonists, antagonists and antibodies can also be used to promote wound  
CC healing, maintain organs before transplantation, and support cell culture  
CC of primary tissues. AA020342-AA020666 represent human secreted protein  
CC amino acid sequences, and related sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO  
CC at: [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 169 AA:

SO Query Match 7.9%; Score 79; DB 22; Length 169;

Best Local Similarity 23.4%; Pred. No. 0.95; Mismatches 50; Indels 68; Gaps 8;

Matches 41; Conservative 16; Indels 68; Gaps 8;

OY 33 DWMDYVVEPPNLEVEFEESTGQNLVKMLENCLSKSKOTKLGCSVLYPEKLTQRIADQV 92  
DB 7 DYM-----LYSRGVC-----RTKST-----DNETSRRNEEV 33

OY 93 LRLSST-EPCGLRGCVMHVNLIEINVCKRLD-----RIVCDSSVPTF 134  
DB 34 MTHSGLMRTCCLEGA-----FRGVCKIDHFPEDADYEODTAEYLLAAVRASSVFPIL 86

OY 135 ELTLVFKQENCSWTSFRDFEFSRGRFSSGFRRTLLSSGFRLYKKKLYSLIGTTV 189  
DB 87 SVTLTFPGGLCYAAS--EFHRS-----RHNVILSAGIFPVSAGLSNILIGIIV 131

RESULT 15

AAR53277 ID AAR53277 standard; Protein: 261 AA.

XX AAR53277;

DT 23-JUN-1994 (first entry)

XX Ced-4 (Q262X).

KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;  
KW embryogenesis; cell death; hydrophilic; transmembrane; region;  
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;  
KW protein synthesis; null phenotype; calcium-binding domain.

XX Caenorhabditis elegans.

XX MO9325685-A.

XX 23-DEC-1993.

XX 14-JUN-1993; 93WO-US05701.

XX 12-JUN-1992; 92US-0897788.

XX 20-NOV-1992; 92US-0979638.

XX (MAST ) MASSACHUSETTS INST TECHNOLOGY.

XX Horvitz HR, Shaham S, Yuan J;

XX WPI: 1994-007542/01.

XX N-PSDB: AA064641.

XX Isolated C elegans cell death genes ced-3 and ced-4 - used to  
XX develop agents to increase or prevent cell death in organisms

XX Claim 10; Fig 1; 127pp; English.

XX The sequences given in AAR53274-78 are encoded by mutant versions of the  
CC C. elegans ced-4 gene. A 2.2 kb mRNA was identified as the wild type  
CC ced-4 transcript. This transcript was present at normal levels in a  
CC ced-3 mutant, suggesting that ced-3 is not a transcriptional regulator  
CC of ced-4. The 2.2 kb transcript is primarily expressed during embryo-  
CC genesis, which is consistent with observed cell deaths in C. elegans.  
CC The wild type Ced-4 protein is 549 amino acids in length and has a  
CC relative molecular mass of 62,877. Ced-4 is highly hydrophilic with

CC a predicted pI of 5.12. There are no obvious transmembrane regions.  
CC The longest hydrophobic region is a segment of 12 amino acids from  
CC residue 382-393. Of the eight mutations which occur within the ced-4  
CC gene one results in a single amino acid substitution and the other  
CC seven appear to prevent either ced-4 RNA splicing or completion of  
CC Ced-4 protein synthesis. These seven mutations establish the null  
CC phenotype of the ced-4 gene, confirming that ced-4 function is not  
CC essential for viability. Ced-4 contains two regions which have  
CC similarity with known calcium-binding domains, suggesting that Ced-4  
CC activity may be modulated by calcium.

XX Sequence 261 AA:

SO Query Match 7.9%; Score 79; DB 15; Length 261;

Best Local Similarity 24.7%; Pred. No. 1.7; Mismatches 46; Indels 68; Gaps 11;

OY 10 KNPASISELD-CGYHPESLSDF--DYMDYVVEPPNLE--VFEESTGQNLVKMLENC 64  
DB 63 KQASELGLIDFPNTNNSHLDLFEDYIDFAINEPDLRPVIAQPSRQML----- 115

OY 65 LSKSKOTKLGCSKVLVPEKLTQRIADQVLRSLSTEPGCLRGCVMHVNLIEINVCKRLDRI 124  
DB 116 ---DKRLLG---NVPKQMT-----CYIR-----EYHVDVIRKLD-- 145

OY 125 VCDSSVPTFELTLVFKQENCSWTSFRDFEFSRGRFSSGFRRTLLSSGFRLYKKKLYSL 184  
DB 146 -----EMCDLDSF--FLFLHGRAGSG--KSVASQAL-----SKSDQL 179

OY 185 IG 186  
DB 180 IG 181

Search completed: October 11, 2002, 14:11:14  
Job time : 64 secs

